Sensitivity analysis of bias of estimates from web surveys with nonrandomized panel selection.

Vladislav Beresovsky[*]

[*]National Center for Health Statistics, 3311 Toledo Rd, Hyattsville, MD 20782

Abstract
Rising costs, growing non-response and uncertain frame coverage stimulate growing interest in the use of web surveys instead of representative samples for estimation of general population characteristics. Previous research has demonstrated that weight adjustment by propensity score constructed on demographic variables may be insufficient for balancing volunteer panels of web surveys selected in nonrandomized fashion. This may result in biased estimates of national characteristics. Assuming systematic dependence of propensity score and target variable on unobserved variable U, we derive dependence of bias on the model parameters associated with U. Asymptotic dependence of relative bias on small values of model parameters associated with U was calculated using Taylor series expansions. Plotted for a wide range of these parameters, relative bias was shown to vary between -1 and 1. Better understanding of the reasons for biased estimates will stimulate development of methods for bias reduction.

Key Words: web panel, reference survey, biased estimates, nonrandomized experiment, propensity score.

Introduction

In recent years there has been an increase of interest in using web surveys for estimating general population characteristics. This innovative method of data collection helps avoid the complications associated with traditional probability sampling and provides for cost reduction and quick turnout, but could result in biased estimates. The reason for bias is the impossibility of defining weights which would correctly project non-probability sample data to the population of interest.

Similar problems occurred in the analysis of data from observational studies, where different medical treatments were non-randomly applied to participating patients. Straightforward estimation of treatment effect was impossible because groups of patients receiving different treatments could differ systematically. Rosenbaum and Rubin (1983b) in their seminal paper proposed estimating treatment effect *conditionally* on probability for a patient to be assigned to one of the groups. This probability, called the *propensity score*, was modeled on patient characteristics available for the complete population. Conditioning on the propensity score removed the difference between patients in different treatment groups, allowing the treatment effect from each group to be projected to the rest of the population.

A justification for using web panels for producing national estimates relies on a similar idea of projecting estimates from web panel data to the remaining population conditionally on the propensity score, calculated as the probability of being included in the web panel. This projection technique is implemented using weights defined as the inverse of the propensity score. It's fair to say that such an estimator of national characteristics from web panels is closely associated with two well-known estimation techniques: (1) estimating the treatment

The findings and conclusions in this paper are those of the author and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

effect from observational study data using propensity score and (2) the Horwitz-Tompson estimator from random sample data.
The major problem with any estimation relying on propensity scores lies in the possibility of bias if insufficient information is available for correctly modeling the propensity score. In the case of observational studies, usually there is a lot of information available for a limited number of involved patients. If participating doctors recorded all of the information they used for making decisions on treatment assignment for each patient, one might assume that the propensity score utilizing all of this information would provide for unbiased estimation of treatment effect. At the same time, the probability of being on a web panel is usually modeled for the general population, for which very little is known besides basic demographic characteristics. In this case it's very likely that important covariates are missing from the propensity score model.
In another paper published in the same year Rosenbaum and Rubin (1983a) proposed a parametric method for assessing sensitivity of estimating treatment effect to an unobserved binary patient characteristic relevant to both treatment assignment and response. Applied to observational study data, this method estimates treatment effect as a function of model parameters associated with the aforementioned unobserved patient characteristic.
In this paper we investigate the effect of ignoring an unobserved covariate in a propensity score model on estimates from a web panel data. In Section 1 a general expression utilizing propensity score is derived for calculating the expectation of an outcome variable using the distribution of a web panel data. An analogy is demonstrated between this expectation and the Horwitz-Tompson estimator from stratified random samples. Assuming parametric distributions for the outcome variable Y and web panel inclusion indicator Z depending on observed and unobserved covariates (X, U), in Section 2 we calculate the relative bias of the expectation of the outcome variable resulting from ignoring unobserved covariate U in the propensity score model. Section 2.2 shows an asymptotic formula for relative bias for small values of model parameters associated with an unobserved covariate. Relative bias in a wide range of model parameters is plotted in Section 2.3. In the Conclusion we summarize the findings of this paper and propose a strategy for minimizing bias of propensity score estimates from web panel data. We advance arguments for the combined use of web surveys and reference randomized surveys to expand data collection while reducing bias of estimates from non-randomized samples.

1. Bias of estimates due to an unobserved covariate

1.1 Justification for using propensity score for unbiased estimation from web panel data

Suppose that we want to estimate the expectation of the random variable of interest Y over the general population $E(Y)$. But data available for estimation comes from a web panel, which is a non-randomly selected group of respondents. Let binomial random variable Z represent an indicator for a person to be included in the web panel $(Z = 1)$. The distribution of Y for people on the web panel could be very different from the general population. In a statistical sense this means that variables Y and Z are correlated. Using Bayes theorem we can express the distribution of Y over the general population through the conditional distributions of Y and Z. Assuming $Z = 1$ allows us to focus on the web panel data:

$$P(Y) = \frac{P(Y|Z=1)\,P(Z=1)}{P(Z=1|Y)} \tag{1.1}$$

If an additional covariate X is observed for the web panel data, the expression above could be considered conditional on this covariate:

$$P\left(Y|X\right) = P\left(Z=1|X\right)\frac{P\left(Y|Z=1,X\right)}{P\left(Z=1|Y,X\right)} \tag{1.2}$$

Finding the expectation of the outcome variable Y using distributions conditional on X involves taking the expectation over the distribution of covariate X. While expectation of the left hand side of (1.2) is taken over the distribution of X in the general population, we want to take the expectation of the right hand side over the distribution of X for the web panel, namely $P(X|Z=1)$. This can be achieved by using Bayes theorem once again, this time for variables Z and X:

$$P\left(X\right) = \frac{P\left(X|Z=1\right)P\left(Z=1\right)}{P\left(Z=1|X\right)} \tag{1.3}$$

From this expression it follows that:

$$E_X\left[g(X)\right] = P\left(Z=1\right)E_{X|Z=1}\left[\frac{g(X)}{P\left(Z=1|X\right)}\right] \tag{1.4}$$

where $g(X)$ is any function of X. With (1.4) in mind, expectations of both sides of (1.2) result in:

$$E_Y\left[Y\right] = E_X\left[E_Y\left[Y|X\right]\right] = P\left(Z=1\right)E_{X|Z=1}\left[E_{Y|X,Z=1}\left[\frac{Y}{P\left(Z=1|Y,X\right)}\right]\right] \tag{1.5}$$

Here, the expectation of Y over the *general population* is expressed through the expectation over *web panel data*. However, (1.5) is not practically applicable because the dependence of inclusion indicator $P\left(Z=1|Y,X\right)$ on Y is unknown. At this moment, we make the *crucial assumption* that, conditionally on X, web panel inclusion indicator Z is independent of target variable Y:

$$P\left(Z=1|Y,X\right) = P\left(Z=1|X\right) \tag{1.6}$$

If indeed all systematic dependences of $P\left(Z=1|Y,X\right)$ were encapsulated in X, then substituting $P\left(Z=1|X\right)$ in (1.5) would produce an unbiased expectation of Y:

$$E_Y\left[Y\right] = P\left(Z=1\right)E_{X|Z=1}\left[\frac{E_{Y|Z=1,X}\left[Y\right]}{P\left(Z=1|X\right)}\right] \tag{1.7}$$

Conditional probability $P(Z=1|X)$ is called the *propensity score*.

1.2 Analogy between propensity score estimator and Horwitz-Thompson estimator for random samples

Expression (1.7) is in many ways similar to the Horwitz-Thomson estimator of the mean from random samples:

$$\hat{y}^{HT} = \frac{1}{N}\sum_{i\in s}\frac{y_i}{\pi_i} \tag{1.8}$$

where π_i are sample inclusion probabilities and N is population size. In both cases, we calculate a weighted expectation (mean for HT estimator) over web panel (or random sample) data to estimate expectation (mean) over the general population. The stochastic mechanism of selecting a web panel can be compared to a Poisson sampling scheme, where each sample element is selected *independently* with probability π_i *assigned* by the sampler's decision. Web panel elements can be also considered independently selected with probability *predefined* by propensity score (1.6). The most obvious analogy exists between (1.7) and the Horwitz-Thompson estimator for stratified simple random samples:

$$\hat{y}^{str} = \frac{1}{N} \sum_{X=x_i} \frac{N_i}{n_i} \sum_{j \in x_i} y_{ij} \tag{1.9}$$

where N_i is the population count for elements with $X = x_i$ and n_i is the corresponding sample count. The propensity score estimator (1.7) transforms exactly into $\hat{y}^{str}$ assuming the following estimators for expectations and probabilities:

- $E_Y\left[Y|Z=1,X\right] = \frac{1}{n_i}\sum_{j \in x_i} y_{ij}$;
- $E_{X|Z=1}[g(X)] = \sum_{X=x_i} \frac{n_i}{n} g(x_i)$, where n is web panel size and $g(X)$ is any function of X;
- $P\left(Z=1|X=x_i\right) = n_i/N_i$;
- $P\left(Z=1\right) = n/N$.

1.3 Expectations under model with an unobserved covariate

Expectation with propensity score (1.7) is unbiased only if conditioning on observed covariate X provides for independence of Y and Z. Let's assume that information in the *observed* covariate X is insufficient to satisfy (1.6), and X must be complemented with the *unobserved* covariate U:

$$P\left(Z=1|Y,X,U\right) = P\left(Z=1|X,U\right) \tag{1.10}$$

Expectation (1.7) becomes:

$$E_Y\left[Y\right] = P\left(Z=1\right) E_{X,U|Z=1}\left[\frac{E_{Y|Z=1,X,U}\left[Y\right]}{P\left(Z=1|X,U\right)}\right] \tag{1.11}$$

To be more specific, suppose that covariate X is a multinomial variable, taking values $x_i,\ i = (1 \ldots I)$ with probabilities $P\left(X=x_i|Z=1\right)$ for the web panel data. The unobserved covariate U is assumed to be binomial with conditional probability distribution $P(U|Z=1,X)$. Then the unbiased expectation of the outcome variable Y is:

$$y^{ub} = P\left(Z=1\right) \sum_{i=1}^{I} P\left(x_i|Z=1\right) \sum_{u=0}^{1} P(u|Z=1,x_i) \frac{E\left(Y|Z=1,x_i,u\right)}{P\left(Z=1|x_i,u\right)} \tag{1.12}$$

This expression utilizes a propensity score modeled on the unobserved covariate U making it impossible to estimate from available data. We may consider estimation using the *observed* marginal propensity score independent of U:

$$P^{obs}(Z=1|X) = \sum_{u=0}^{1} P(U|X)P(Z=1|X,U) \tag{1.13}$$

Here, the distribution of the unobserved covariate U conditional on X is defined on the general population, while in (1.12) it's defined on the web panel data. Using the observed propensity score will result in biased expectation of Y:

$$y^b = P\left(Z=1\right)\sum_{i=1}^{I}\frac{P\left(x_i|Z=1\right)}{P^{obs}\left(Z=1|x_i\right)}\sum_{u=0}^{1}P(u|Z=1,x_i)E\left(Y|Z=1,x_i,u\right) \quad (1.14)$$

To summarize, expression (1.12) is unbiased but impossible to estimate from the observed data. Expression (1.14) is generally biased, but it can be estimated in practice. The goal of this paper is to calculate relative bias $relBias(y^b) = \left(y^b - y^{ub}\right)/y^{ub}$ assuming specific parametric models for an outcome variable Y and propensity score. Studying analytic dependence of $relBias(y^b)$ on model parameters associated with an unobserved covariate U will provide information about the significance of this bias in various situations.

2. Parametric expression for relative bias

2.1 Parametric dependence on an observed and unobserved covariates

Let's use the following notations for distributions of an observed multinomial covariate X and unobserved binomial covariate U:

$$P\left(X=x_i|Z=1\right)=\phi_i, \text{where } i\in(1\ldots I) \text{ and } \sum_{i=1}^{I}\phi_i=1 \quad (2.1a)$$

$$P\left(U=u|X=x_i\right)=P\left(U=u|Z=1,X=x_i\right)=\varphi_i(u), \quad (2.1b)$$

$$\text{where } u\in(0,1) \text{ and } \sum_{u=0}^{1}\varphi_i(u)=1$$

From here we will consider the conditional distribution $P\left(U|X\right)$ to be the same for a web panel data $(Z=1)$ and the general population. We have found that this assumption does not qualitatively alter the main conclusions of our study.

For a binary outcome variable Y and web panel inclusion indicator Z we assume Bernoulli distributions conditional on (X,U) with *logit* link function:

$$P\left(Y=1|Z=1,X,U\right)=\frac{exp(\gamma_i+\delta_i u)}{1+exp(\gamma_i+\delta_i u)} \quad (2.2a)$$

$$P\left(Z=1|X,U\right)=\frac{exp(\alpha_i+\beta_i u)}{1+exp(\alpha_i+\beta_i u)} \quad (2.2b)$$

Note that the distribution of Y is defined over the web panel data and the distribution of Z is over the general population.

The model described above allows us to write parametric expressions for both unbiased (1.12) and biased (1.14) expectations of the outcome variable Y:

$$y^{ub(b)}=\sum_{i=1}^{I}\phi_i A_i^{ub(b)}, \quad (2.3a)$$

$$A_i^{ub}=\sum_{u=0}^{1}\varphi_i(u)\left(\frac{exp(\gamma_i+\delta_i u)}{1+exp(\gamma_i+\delta_i u)}\right)\left(\frac{exp(\alpha_i+\beta_i u)}{1+exp(\alpha_i+\beta_i u)}\right)^{-1} \quad (2.3b)$$

$$A_i^{b}=\left(\sum_{u=0}^{1}\varphi_i(u)\frac{exp(\gamma_i+\delta_i u)}{1+exp(\gamma_i+\delta_i u)}\right)\left(\sum_{u=0}^{1}\varphi_i(u)\frac{exp(\alpha_i+\beta_i u)}{1+exp(\alpha_i+\beta_i u)}\right)^{-1} \quad (2.3c)$$

In this notation relative bias due to ignoring an unobserved covariate U is:

$$relBias\left(y^b\right) = \frac{y^b - y^{ub}}{y^{ub}} = \frac{\sum_{i=1}^{I} \phi_i A_i^b}{\sum_{i=1}^{I} \phi_i A_i^{ub}} - 1 \tag{2.4}$$

Analysis of this expression can be simplified by considering ratio $a_i = A_i^b / A_i^{ub} - 1$ for a specific value x_i of the observed multinomial covariate X. Then summation (2.4) becomes the weighted sum of a_i:

$$relBias\left(y^b\right) = \frac{\sum_{i=1}^{I} \phi_i A_i^{ub} a_i}{\sum_{i=1}^{I} \phi_i A_i^{ub}} \tag{2.5}$$

Analysis of this weighted sum can be effectively reduced to analysis of individual a_i.

2.2 Relative bias under small effect of an unobserved covariate

Relative bias depends on parameters $(\varphi_i, \beta_i, \delta_i)$ associated with an unobserved covariate U. Below we identify values of these parameters for which $a_i = 0$ in (2.5) and give practical interpretations of these conditions.
First, consider $\beta_i = 0$. This corresponds to a negligible effect of unobserved covariate U on propensity score (2.2b) for a given x_i and zero contribution a_i to $relBias\left(y^b\right)$.
Second, suppose that unobserved covariate U is homogeneously distributed $\varphi_i(0) = 0$ or $\varphi_i(1) = 0$ for all data elements with a particular x_i. In this case also $a_i = 0$. It is reasonable to assume that bias will grow with heterogeneity of U, achieving maximum value for $\varphi_i(0) = \varphi_i(1) = 0.5$.
Conditional expectation of an outcome variable on *population* $E(Y|U, X)$ is the expression in squared brackets in (1.11). Its independence of an unobserved covariate imposes the following relationship between model parameters:

$$\left(\frac{exp(\gamma_i + \delta_i)}{1 + exp(\gamma_i + \delta_i)}\right)\left(\frac{exp(\alpha_i + \beta_i)}{1 + exp(\alpha_i + \beta_i)}\right)^{-1} = \left(\frac{exp(\gamma_i)}{1 + exp(\gamma_i)}\right)\left(\frac{exp(\alpha_i)}{1 + exp(\alpha_i)}\right)^{-1} \tag{2.6}$$

This also provides for $a_i = 0$ and zero relative bias (2.5). By rearranging terms in this expression, it can be shown that relative bias is small when conditional expectation of an outcome variable on a *web panel data* and propensity score equally increase or decrease with U. When U has the opposite effect on these distributions, relative bias tends to increase. Further analysis confirms this general conclusion, which could be useful as an indicator of possible large bias for some of the estimates.
These simple findings facilitate obtaining an approximate analytical dependence of relative bias (2.5) on model parameters in situations where dependence of the propensity score on an unobserved covariate is small $\beta_i \ll 1$:

$$\beta_i \ll 1,\ \delta_i \ll 1,\ \varphi_i(0) \ll 1 \tag{2.7a}$$

$$a_i = \frac{\varphi_i(0)\beta_i}{1 + \exp(\alpha_i)}\left(\frac{\delta_i}{1 + \exp(\gamma_i)} - \frac{\beta_i}{1 + \exp(\alpha_i)}\right)$$

$$\beta_i \ll 1,\ \delta_i \ll 1, \varphi_i(0) \approx 0.5 \tag{2.7b}$$

$$a_i = \frac{\beta_i}{4(1 + \exp(\alpha_i))}\left(\frac{\delta_i}{1 + \exp(\gamma_i)} - \frac{\beta_i}{1 + \exp(\alpha_i)}\right)$$

Relative bias is proportional to small deviations of model parameters β_i and $\varphi_i(0)$ from zero values, in agreement with previously obtained conditions for zero bias. The term in brackets may be obtained from imposing a condition of zero derivative of conditional expectation of Y over the general population $dE(Y|U,X)/dU = 0$ for $U = 0, 1$ and small values of β_i and δ_i. Note that a_i is also inversely proportional to $1+\exp(\alpha_i)$, which means that bias is potentially larger for web panel elements with small observed propensity score (1.13) .

2.3 Relative bias in a wide range of model parameters

The main reasons for biased estimates were explicitly demonstrated for small effects of an unobserved covariate in the analytic expressions (2.7a-b). At the same time, expressions (2.3a-c) allow us to calculate the contribution a_i to relative bias for a given x_i (2.5) for all values of model parameters. Dependence of a_i on parameter β_i is presented in Figure 1.

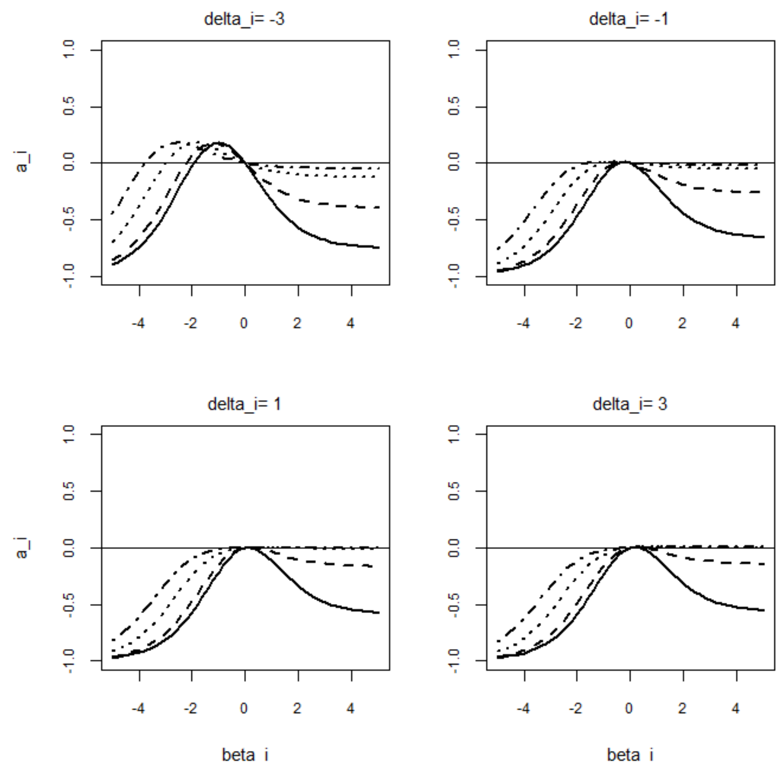

Figure 1: Relative bias a_i for a given x_i depending on parameter β_i. Parameter δ_i is set to $(-3, -1, 1, 3)$. Different values of parameter $\alpha_i = (-2, -0.5, 1, 2)$ are marked by solid, dashed, dotted and dot-dashed line types. Parameters $\gamma_i = 1$ and $\varphi_i(0) = \varphi_i(1) = 0.5$ were kept constant in all cases.

As expected, relative bias $a_i = 0$ for $\beta_i = 0$. It changes between 1 and -1 for all other values of β_i. In most cases relative bias is negative. It becomes significantly positive only for large and negative values of δ_i. It can be proven analytically that $a_i \to 1$ for moderate α_i and γ_i when both $\beta_i, \delta_i \to -\infty$.
The absolute value of negative relative bias always increases when parameter α_i decreases and becomes negative. This means a larger negative contribution to bias from web panel elements with small "observed" propensity scores (1.13). For example, for any δ_i in the whole region $\beta_i > 0$ relative bias is very small for $\alpha_i > 0$ but becomes substantially negative for $\alpha_i < 0$. The situation is similar for $\beta_i < 0$, when both β and δ_i are not too

negative.
Asymptotic expressions for a_i in the limit $\beta_i \to \pm\infty$ can be easily obtained from the general formula (2.3a-c):

$$\beta \to \infty; \tag{2.8a}$$
$$a_i \to \frac{2d\exp(\alpha_i)(1+\exp(\alpha_i))}{(1+d\exp(\alpha_i))(1+2\exp(\alpha_i))} - 1, \text{where}$$
$$d = 1 + \exp(\delta_i)(1 + 2\exp(\gamma_i))$$
$$\beta \to -\infty; \tag{2.8b}$$
$$a_i \to [\exp(\delta_i)(1+\exp(\gamma_i))]^{-1} - 1$$

For $\beta \to \infty$ relative bias is driven by the value of $\exp(\alpha_i)$. Large positive α_i corresponds to almost unbiased expectation $a_i \to 0$. As α_i becomes more negative, relative bias grows in a negative direction.
In the limit $\beta_i \to -\infty$, relative bias strongly depends on $\exp(\delta_i)$. Expression (2.8b) shows that $a_i \to -1$ for all α_i when $\exp(\delta_i) \gg 1$. In the opposite limit $\delta_i \to -\infty$, relative bias $a_i \to 1$ for all values of α_i. This unrealistic scenario was not plotted in Figure 1. For moderately negative δ_i relative bias a_i goes through the positive maximum and then becomes negative again as β_i further decreases.
The dependence of relative bias on model parameters presented in Figure 1 was calculated for maximum heterogeneity $\varphi_i(0) = 0.5$ of an unobserved covariate U. Relative bias is expected to decrease proportionally (2.7a-b) when $\varphi_i(0)$ becomes smaller.

Conclusions and practical implications

In Section 1, we derived a general expression (1.5) for the expectation $E_Y[Y]$ of target variables over *population data* utilizing the expectation of Y over *web panel data* divided by web panel inclusion probability $P(Z=1|Y,X)$. Because the inclusion probability depends on outcome variable Y, it cannot be calculated in practice. However, if auxiliary covariates X closely correlate with Y, the dependence of the web panel inclusion probability on Y can be reduced and ultimately ignored (1.6). Such approximation reduces the "true" inclusion probability to propensity score $P(Z=1|X)$, which allows estimating $E_Y[Y]$ from web panel data (1.7). Bias of this approximation depends on how closely inclusion probability $P(Z=1|X,Y)$ can be approximated by propensity score $P(Z=1|X)$.
In Section 2.2 we demonstrated that when the effect of an unobserved covariate on the propensity score (2.2b) is small ($\beta_i \ll 1$), the resulting relative bias is also small and proportional to β_i (2.7a-b). It was also shown to be small when unobserved covariate U is homogeneously distributed $\varphi_i(0) \ll 1$ for a given observed covariate x_i and when the population distribution of outcome variable Y is independent of U.
However, relative bias may substantially deviate from zero when the effect of an unobserved covariate on the propensity score is significant. Dependence of relative bias a_i for a given observed variable x_i in a wide range of model parameters was plotted in Figure 1.
Results of this paper suggest that bias of estimates from a web panel data might be small when observed covariate(s) X used for modeling the propensity score are strongly correlated with an outcome variable Y. For example, in the context of health surveys this means that using only demographic covariates for calculating the propensity score may be insufficient, because it's easy to imagine two groups of people who are demographically identical but who have different health characteristics and propensity to join a web panel. These differences may be due to characteristics other than demographic ones, such as cultural and

social habits, genetic traits, etc.

Common life experience teaches us that generally one cannot count on a “free lunch”. Planners of a probability survey may rely on stratification by demographic variables alone because they apply the random mechanism of sample selection within predefined strata. However, this method involves significant effort and expense related to correct definition of survey frame, random sample selection and data collection activities.

When web panels are used for estimation of general population characteristics, even selecting a demographically “representative” web panel does not provide a guarantee of randomness, because people who agree to join a web panel may be substantially different from the rest of the population. The solution in this case is to accumulate as much information as possible about both the web panelists and the general population, which could be used for modeling the propensity score. Importantly, this information must be closely correlated with the outcome variables to be estimated from the web panel data.

It’s possible to work systematically to reduce bias of estimates from web panel data when a web panel is used as a *supplement* to a regular probability survey, often referred to as a *reference* survey. Survey planners must decide in advance how they are going to spread questions between the web panel and the reference survey. There must be a carefully selected limited group of questions Y_s, shared by both the reference survey and the web panel. These questions must be potentially strongly correlated with questions Y_w slated to the web panel only. Presuming a strong correlation between Y_w and Y_s, we can estimate the web panel inclusion probability $P\left(Z=1|X,Y_w\right)$ by the propensity score $P_s\left(Z=1|X,Y_s\right)$. Propensity score P_s can be estimated from data collected from the web panel and the reference survey. Resulting bias of estimates will depend on the degree of correlation between Y_w and Y_s.

The list of shared variables Y_s must be informative enough to substitute for Y_w, but it does not have to be very extensive. Our research suggests that it does not make sense to have shared questions which assume a high degree of homogeneity of responses from certain demographic groups. For example, it does not make sense to ask young people if they have a chronic cancer condition, but it does make sense to ask this question of people above 65 years of age.

The statistical findings of this paper suggest that survey planners can use a combination of web panels and reference probability surveys to work systematically on reducing bias of estimates from web panels. Further research is needed to refine the implementation of this methodology.

References

Rosenbaum, P. R. and Rubin, D. B. (1983a). Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. *J.R. Statist.Soc. B*, 45:212–218.

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