Simple Approaches to Estimating the Variance of the Propensity Score Weighted Estimator Applied on Volunteer Panel Web Survey Data – a Comparative Study

Annica Isaksson[1], Sunghee Lee[2]
Statistics Sweden[1]
UCLA Center for Health Policy Research[2]

Abstract

Survey data collection from volunteer Web panels is growing in popularity. While this practice has obvious advantages in terms of cost, flexibility, and speed, estimates from such data for characteristics of the general population may be seriously biased due to low web penetration in the general population, nonprobability samples, and high nonresponse rates. The propensity score adjustment was introduced as a possible solution to the selection bias in observational studies. Its modification, propensity score weighting using a reference survey, has been suggested as a remedy of problems in volunteer panel web surveys. We can derive a propensity score weighted estimator that is less biased, if not unbiased, of the parameter of interest under certain model assumptions. However, there is no evident solution for its variance estimation. In the present study, simple variance estimators are compared: (1) the estimator applicable under poststratification, (2) a model-based estimator, and (3) the jackknife estimator. The study focuses on the performance of the estimators in a simulation designed from a real web survey.

Keywords: Non-probability sampling, Variance estimation, Simulation.

1. Introduction

Propensity score adjustment has been introduced as a possible solution of selection bias in observational studies (Rosenbaum and Rubin 1983, 1984, D'Agostino and Rubin 2000). Its modification, propensity score weighting using a reference survey, has been suggested as a remedy of problems in volunteer panel Web surveys (Terhanian et al. 2001). The propensity score weighted estimator is less biased, if not unbiased, of the parameter of interest, under certain model assumptions. A remaining problem, however, is to estimate its variance. In this paper, some variance estimation approaches are compared through a simulation, designed to imitate a real application. It is based on data from the 2003 Michigan Behavioral Risk Factor Surveillance System (BRFSS), and Web samples are drawn with the same allocation over key demographic variables as respondents in a Harris Interactive Web survey. As measures of comparison we calculate the relative biases of the different variance estimators, as well as the coverage rate associated with each of them.

The propensity score adjusted point estimator is introduced in sec. 2 and the different variance estimators in sec. 3. In sec. 4, the simulation design is presented and the simulation results analyzed. Finally, in sec. 5, some conclusions are drawn.

2. A Propensity Score Estimator

Let y_k define a random study variable associated with individual k included in the general population U of size N (the actual y_k is taken as a realization of this random variable). The variables $y_1,...,y_N$ are regarded as independently and identically distributed with a common mean μ. According to general properties of a random sample, the expectation of the mean $\bar{y}_U = \sum_U y_k / N$ is then equal to μ. The inference problem is to estimate μ from a survey relying on two samples: a nonrandom Web sample and a probability sample. The Web sample, s_W, constitutes the greater part of the total sample, while the probability sample, s_P, is small and observed for the sole purpose of aiding estimation. The sample s_P is selected by simple random sampling (SI) from U, whereas s_W is selected from a volunteer Web panel, $U_W \left(U_w \subset U \right)$. Whether individual $k \in U$ is included in the Web panel or not is indicated by the random variable z_k: $z_k = 1$ if $k \in U_w$, else $z_k = 0$. The sizes of s_W, s_P and U_W are denoted n_W, n_P and N_W, respectively. The size of the total sample $s = s_W \cup s_P$ is $n = n_w + n_P$.

We restrict attention to the following estimation approach. Let $\mathbf{x}_k$ denote the value of a vector of covariates $\mathbf{x}$ associated with individual $k \in U$. The (unknown) propensity score of individual k is defined

as $e(\mathbf{x}_k) = \Pr(z_k = 1 | \mathbf{x}_k)$. A regression model for $e(\mathbf{x}_k)$ as a function of $\mathbf{x}_k$ is formulated and used to estimate the propensity scores from s. Then, s is divided into H classes, $s_1, \ldots, s_h, \ldots, s_H$ of sizes $n_1, \ldots, n_h, \ldots, n_H$, by similar estimated propensity scores (based on Cochran, 1968, the number of classes is often set to five.) The aim of the stratification is to create classes within which y_k and z_k are independent. For $h = 1, \ldots, H$, the intersections $s_{Wh} = s_W \cap s_h$ of size n_{Wh} and $s_{Ph} = s_P \cap s_h$ of size n_{Ph} are defined, as well as the corresponding vectors $n_W = (n_{W1}, \ldots, n_{Wh}, \ldots, n_{WH})$ and $n_P = (n_{P1}, \ldots, n_{Ph}, \ldots, n_{PH})$. Let $d_h = n_{Ph}/n_P$ and $\bar{y}_{s_{Wh}} = \sum_{s_{Wh}} y_k / n_{Wh}$. A propensity score weighted estimator of μ, similar to the weighting cell estimator in Oh and Scheuren (1983, eqn. (21)) designed for nonresponse adjustment, is given by

$$\hat{\mu}_{PS} = \frac{\sum_{h=1}^{H} \sum_{s_{Wh}} \phi_{hk}^{-1} y_k}{\sum_{h=1}^{H} \sum_{s_W} \phi_{hk}^{-1}} = \sum_{h=1}^{H} d_h \bar{y}_{s_{Wh}} \qquad (1)$$

where each inverse phi is the product of a 'base weight' b_k and an 'adjustment weight' a_h:

$$\phi_{hk}^{-1} = b_k a_h = \frac{N}{n_W} \frac{n_{Ph} n_W}{n_P n_{Wh}} = \phi_h^{-1}$$

(Lee, 2004, sec. 6.2). Note that the base weight would be a proper design weight if s_W had been an SI sample from U.

Although our paper is strictly concerned with the estimator in eqn. (1), alternative uses of propensity scores certainly exist. In particular, as suggested by Cassel et al. (1983) for nonresponse adjustment, inverses of estimated propensity scores can be used to weight observations directly. The propensity score estimator used by Beissel-Durrant and Skinner (2004) for measurement error adjustment is constructed in this manner. Pros and cons of direct weighting by inverse estimated propensity scores (for nonresponse adjustment) are discussed in Little (1986, p. 146) and Little and Rubin (1987, p. 58). Their objections to the approach include that it may produce an estimator with extremely high variance and with heavy reliance on the model assumptions.

3. Variance Estimation Approaches

In this section, some different approaches to estimation of the variance of $\hat{\mu}_{PS}$ are presented. Later, in section 4, their behavior will be studied in a simulation.

The variance estimator treated in section 3.1 is used in finite population sampling, whereas the one treated in section 3.2 is derived through statistical modeling. Both are fairly easy to compute. In section 3.3, we consider variance estimation by a re-sampling method, the jackknife (JK) method, which computationally is much more demanding.

3.1 Poststratification

A straight-forward way of estimating the variance of $\hat{\mu}_{PS}$ is the following. As in finite population sampling, look upon the y_k's as fixed values. Also view the stratification by propensity scores as a way of identifying fix strata (of fix but unknown sizes $N_1, \ldots, N_h$) in the population. In the estimation, use the adjusted first and second inclusion probabilities

$$\pi_{hk}^* = \phi_{hk} = \frac{n_W}{Na_h} = \pi_h^*; k = l \in s_W$$

$$\pi_{hkl}^* = \phi_{hkl} = \frac{n_W(n_W - 1)}{Na_h(Na_h - 1)}; k \neq l \in s_W \qquad (2)$$

Use of Result 7.6.1 in Särndal et al. (1992) together with π_{hk}^* and π_{hkl}^* produces a poststratified point estimator of a finite population mean:

$$\hat{\bar{y}}_U = \sum_{h=1}^{H} \frac{N_h}{N} \bar{y}_{s_{Wh}} \qquad (3)$$

and a corresponding variance estimator

$$\hat{V}(\hat{\bar{y}}_U) = \sum_{h=1}^{H} \left(1 - \frac{n_W}{Na_h}\right)\left(\frac{N_h}{N}\right)^2 \frac{s_{Wh}^2}{n_{Wh}} \qquad (4)$$

where $s_{Wh}^2 = \sum_{s_{Wh}} (y_k - \bar{y}_{s_{Wh}})^2 / (n_{Wh} - 1)$. Derivations of eqn. (3)-(4) are given in Appendix. The stratum proportions N_h/N in the formulas are unknown but may be estimated by d_h. Then, $\hat{\bar{y}}_U = \hat{\mu}_{PS}$. As a small simplification, we omit the finite population correction in (4) and restrict attention to the variance estimator

$$\hat{V}_{POST} = \sum_{h=1}^{H} d_h^2 \frac{s_{Wh}^2}{n_{Wh}} \qquad (5)$$

Eqn. (5) would simplify further if s was divided into strata of equal sizes in terms of n_{Ph}.

3.2 A Model-Based Approach

In Isaksson et al. (2004), the y_k's are looked upon as random variables. A variance estimator is derived by use of the following statistical models for y_k and n_P:

- Conditional on $e(x_k)$, the study variable values y_k for $k \in s_h$, h=1,...,H, are iid random variables with expectation μ_h variance σ_h^2.

- Each individual $k \in s_P$ is independently assigned membership in one of H classes. For each individual, the probability of being assigned to class h is D_h. Thus, n_P has a multinomial distribution with n_P trials, H possible outcomes, and cell probabilities $D_1,...,D_h,...,D_H$.

It is further assumed that $\bar{y}_{s_{wh}}$ and d_h (h=1,...,H) are independent. Note that the statistical properties of n_W are not modeled – instead, all derivations are conditioned on n_W. Derivation of the model variance of $\hat{\mu}_{PS}$, and subsequent replacement of unknown model parameters included in the variance expression with their sample analogues, results in the following variance estimator:

$$\hat{V}_{MOD} = \frac{1}{n_P} \sum_{h=1}^{H} \left[d_h \left(\bar{y}_{s_{wh}} - \bar{y}_s \right)^2 + d_h \left(1 - d_h \right) \frac{s_{Wh}^2}{n_{Wh}} \right]$$
$$+ \sum_{h=1}^{H} d_h^2 \frac{s_{Wh}^2}{n_{Wh}}$$
$$= \frac{1}{n_P} \sum_{h=1}^{H} \left[d_h \left(\bar{y}_{s_{wh}} - \bar{y}_s \right)^2 + d_h \left(1 - d_h \right) \frac{s_{Wh}^2}{n_{Wh}} \right]$$
$$+ \hat{V}_{POST} \qquad (6)$$

Since the first term in eqn. (6) can not be negative, for a given sample, the $\hat{V}_{MOD}$ estimate is always at least as large as the $\hat{V}_{POST}$ estimate.

3.3 Jackknife

An estimation approach of different nature than those treated in sections 3.1-3.2 is the subsampling replication technique. We restrict our attention here to one such technique; the JK. Our application of the method starts with a random division of the 'parent sample' (in our case: the Web sample) into groups. Let g denote the desired number of groups. As described in Wolter (1985, sec. 2.4.1 (i)), the first random group (RG) is obtained by drawing an SI sample of size $m = [n_w / g]$ from s_W; the second RG by drawing an SI sample of size m from the remaining n_W-m units in the parent sample; and so on. (If n_w / g is not an integer, the

excess units are ignored.) From Wolter (1985, eqn. (4.3.5)), the corresponding JK estimator of variance is given by

$$\hat{V}_{JK} = \sum_{\alpha=1}^{g} \frac{g-1}{g} \left(\hat{\mu}_{PS(\alpha)} - \hat{\mu}_{PS} \right)^2 \qquad (7)$$

where $\hat{\mu}_{PS(\alpha)}$ is an estimator of the same functional form as $\hat{\mu}_{PS}$, only with group α of observations omitted.

4. Simulation

How do the variance estimators presented in sec. 3 behave in practice? In order to gain some insights about this, we perform a simulation.

4.1. Design

The population and the sampling procedure presented in this section equal those used in Case Study II in Lee (2004, ch. 7).

Population The simulation is based on data from BRFSS: a survey on risk behaviors linked to diseases and injuries among adults (aged 18 years and older) who live in households. Among the study variables, we find tobacco use, health care coverage, HIV/AIDS knowledge and prevention, physical activity, and fruit and vegetable consumption. The survey is a collaborative project of the Centers for Disease Control and Prevention and the U.S. states and territories. In each participating state, a probability sample of adults in households is selected and observed through computer-assisted telephone interviewing. The collected data are used, among other things, for measuring progress toward state and national health objectives. For additional information on the BRFSS, see its homepage www.cdc.gov/brfss/.

In our application, we restrict attention to BRFSS data from 2003 and the state of Michigan. In this state, respondents are selected by a list-assisted RDD method from strata defined by phone bank density and probability of phone number being listed (Cook and Rafferty, 2004). The original data set contains 3,551 individuals. We use however only the 3,410 cases with complete data on the variables Web access ownership, age, gender, race and education. From this data set, by sampling with replacement, our BRFSS pseudo-population U of size N=20,000 is created.

The subset of the pseudo-population containing individuals who have Web access at home constitute our Web pseudo-population U_W. The size of U_W is N_W =12,674.

Sampling In each simulation round, we draw a reference sample by SI from U, as well as a sample mimicking a real Harris Interactive volunteer Web panel sample [1] from U_W. More precisely, the Web sample is selected by stratified SI from U_W, with stratification by age, gender, education and race, and with the same allocation over strata as the Harris sample (for details on the allocation, see Lee, 2004, table 7.6). Let the reference sample and the Web sample selected in iteration i, $i=1,\ldots,I$, be denoted $s_{P(i)}$ (of size $n_{P(i)}$) and $s_{W(i)}$ (of size $n_{W(i)}$), respectively. In practice, we use $n_{P(i)} = 500$, $n_{W(i)} = 1500$ and I=1000.

For the JK method, in each iteration, the Web sample is randomly divided into g=30 groups. As the number of groups increases, so does the computer time, while the precision of the estimator is expected to improve.

4.2. Estimation

As described in sec. 2, as part of the propensity score estimation approach, a regression model for $e(\mathbf{x}_k)$ is formulated and used to estimate the propensity scores from s. We formulate a standard logistic regression model for $e(\mathbf{x}_k)$ as a function of $\mathbf{x}_k$ (Neter et al. 1996, eqn. 14.37) in which values of z_k $(k \in U)$ are independent Bernoulli random variables having conditional expected values

$$e(\mathbf{x}_k) = \frac{\exp(\beta' \mathbf{x}_k)}{1 + \exp(\beta' \mathbf{x}_k)} \qquad (8)$$

where $\beta' = \lfloor \beta_0, \beta_1, \ldots, \beta_{(p-1)} \rfloor$ and $\mathbf{x}'_k = \lfloor 1, x_{1k}, \ldots, x_{(p-1,k)} \rfloor$. The expected values $e(\mathbf{x}_k)$ are then estimated by

$$\hat{e}(\mathbf{x}_k) = \frac{\exp(b' \mathbf{x}_k)}{1 + \exp(b' \mathbf{x}_k)} \qquad (9)$$

where b is a vector of maximum likelihood estimates of $\beta_0, \beta_1, \ldots, \beta_{(p-1)}$.

It is not obvious which covariates that shall be included in the $\mathbf{x}$-vector. At our disposal, we have 30 covariates (see Lee, 2004, Table 7.7): some continuous, others categorical. Among the 30 are the variables we use for stratification when the Web sample is selected (age, gender, education and race).

---

[1] Obtained via personal communication with Matthias Schonlau.

As in Case Study II in Lee (2004, ch. 7), we try five different compositions of the $\mathbf{x}$-vector:

- M1: Only the stratification variables
- M2: All covariates
- M3: Covariates with p-value not exceeding .2
- M4: All covariates except the stratification variables
- M5: Covariates with p-value not exceeding .2 which are not stratification variables

The estimated propensity scores are used to divide s into classes. More precisely, the sample is sorted in ascending order by estimated propensity score and then divided into five classes of equal size in terms of n_h. The propensity score estimator of μ is then calculated for three different study variables:

- y_1: High blood pressure (1=yes, 0=no). Question wording: "Have you ever been told by a doctor, nurse or other health professional that you have high blood pressure?")
- y_2: Has smoked 100 cigarettes (1=yes, 0=no). Question wording: "Have you smoked at least 100 cigarettes in your entire life?")
- y_3: Does vigorous physical activities (1=yes, 0=no). Question wording: "Now thinking about the vigorous physical activities you [fill in] in a usual week, do you do vigorous activities for at least 10 minutes at a time, such as running aerobics, heavy yard work, or anything else that causes large increase in breathing or hear rate?"

4.3. Simulation Output

For an arbitrary y-variable, and for a given iteration i ($i=1,\ldots,I$), let the mean of the reference sample be denoted $\bar{y}_{s_{P(i)}} = \sum_{s_{P(i)}} y_k / n_{P(i)}$ and the mean of the Web sample be denoted $\bar{y}_{s_{W(i)}} = \sum_{s_{W(i)}} y_k / n_{W(i)}$. The propensity score adjusted estimator of the mean, based on $\mathbf{x}$-vector Mv ($v=1,\ldots,5$), is denoted $\hat{\mu}_{PS,Mv(i)}$. In the simulation, for an arbitrary y-variable, the relative bias RB of $\bar{y}_{s_w}$ is approximated by

$$RB(\bar{y}_{s_w}) = \frac{\bar{\bar{y}}_{s_w} - \bar{y}_U}{\bar{y}_U} \qquad (10)$$

where $\bar{\bar{y}}_{s_w}$ is the average (over I iterations) of the observed Web sample means $\left(\bar{\bar{y}}_{s_w} = \sum_{i=1}^{I} \bar{y}_{s_{W(i)}} / I \right)$. In the same manner, the relative bias of the propensity score

adjusted estimator $\hat{\mu}_{\mathrm{PS,M}v}$ ($v=1,\ldots,5$) is approximated by

$$RB\left(\hat{\mu}_{\mathrm{PS,M}v}\right)=\frac{\overline{\hat{\mu}}_{\mathrm{PS,M}v}-\overline{y}_{U}}{\overline{y}_{U}} \qquad (11)$$

where $\overline{\hat{\mu}}_{\mathrm{PS,M}v}$ is the average (over I iterations) of the observed propensity score adjusted estimates based on x-vector Mv $\left(\overline{\hat{\mu}}_{\mathrm{PS,M}v}=\sum_{i=1}^{I}\hat{\mu}_{\mathrm{PS,M}v(i)}/I\right)$.

Consider the variance estimation approaches presented in sec. 3. Let $\hat{V}_{av}$ denote the variance estimator under approach a (a=POST, MOD, JK) based on x-vector Mv ($v=1,\ldots,5$). In the simulation, the observed value on $\hat{V}_{av}$ for an iteration i ($i=1,\ldots,I$) is denoted $\hat{V}_{av(i)}$. The relative bias of $\hat{V}_{av}$ as estimator of the true variance of $\hat{\mu}_{\mathrm{PS,M}v}$ is approximated by

$$RB\left(\hat{V}_{av}\right)=\frac{\overline{\hat{V}}_{av}-V\left(\hat{\mu}_{\mathrm{PS,M}v}\right)}{V\left(\hat{\mu}_{\mathrm{PS,M}v}\right)} \qquad (12)$$

where $\overline{\hat{\mu}}_{\mathrm{PS,M}v}=\sum_{i=1}^{I}\hat{\mu}_{\mathrm{PS,M}v(i)}/I$; $\overline{\hat{V}}_{av}=\sum_{i=1}^{I}\hat{V}_{av(i)}/I$ and $V\left(\hat{\mu}_{\mathrm{PS,M}v}\right)=\sum_{i=1}^{I}\left(\hat{\mu}_{\mathrm{PS,M}v(i)}-\overline{\hat{\mu}}_{\mathrm{PS,M}v}\right)^{2}/I$.

The coverage rate associated with approach a and x-vector Mv is calculated as

$$CR(a,\mathrm{M}v)=\frac{1}{I}\sum_{i=1}^{I}\lambda_{i} \qquad (13)$$

where

$$\lambda_{i}=\begin{cases}1 & \text{if } \mu_{\mathrm{PS,M}v(i)}-1.96\sqrt{\hat{V}_{av(i)}}<\mu\le\mu_{\mathrm{PS,M}v(i)}+1.96\sqrt{\hat{V}_{av(i)}}\\ 0 & \text{otherwise}\end{cases}$$

In each iteration, 1.96 from the normal distribution table is used for calculating the confidence limits. Thus, our intended coverage rate is 95 per cent.

4.4. Results

Even though this paper is focused on variance estimation, we start by taking a brief look at the bias of the propensity score point estimator. The relative biases of the Web sample mean and of the variations of $\hat{\mu}_{\mathrm{PS}}$ are presented in Table 1. We see that the propensity score

adjustment successfully removes a substantial part of the bias of the unweighted Web sample mean. For different study variables, different choices of x-vectors appear to be preferable from a bias reduction point of view. Generally speaking, however, the vectors M1, M2 and perhaps M3 seem to best at removing bias.

Point estimator	v	Study variable		
		y_1	y_2	y_3
$\overline{y}_{s_W}$	-	-13.7	-10.6	17.6
$\hat{\mu}_{\mathrm{PS}}$	1	**-0.1**	-5.7	8.9
$\hat{\mu}_{\mathrm{PS}}$	2	-4.7	-5.3	5.9
$\hat{\mu}_{\mathrm{PS}}$	3	-10.0	**-5.0**	6.0
$\hat{\mu}_{\mathrm{PS}}$	4	-9.5	-8.6	**5.0**
$\hat{\mu}_{\mathrm{PS}}$	5	-7.8	-8.4	8.8

Table 1: Relative bias in per cent for various combinations of point estimator and x-vector for study variables y_1-y_3. The lowest absolute relative bias for each y-variable is highlighted.

In Table 2, the relative biases of different variance estimators are presented. We see that throughout, the relative bias of the JK estimator is very large. Typically, use of the x-vector M4 or M5 gives the smallest bias. Using the 'best' choice of x-vector, the estimator $\hat{V}_{\mathrm{MOD}}$ produces the lowest absolute relative bias for variable y_1 and y_3; $\hat{V}_{\mathrm{POST}}$ the lowest absolute relative bias for variable y_2.

Study variable	v	a		
		POST	MOD	JK
y_1	1	-16.8	-14.6	78.5
	2	-8.5	-7.6	49.5
	3	-5.4	-4.8	44.3
	4	-2.7	**-1.5**	20.2
	5	**1.6**	3.3	**18.0**
y_2	1	-3.9	-2.4	70.6
	2	-5.2	-4.3	58.2
	3	-4.1	-3.1	46.9
	4	0.9	2.0	28.0
	5	**0.6**	**1.8**	**18.7**
y_3	1	**-4.3**	-2.7	58.6
	2	-10.6	-9.1	48.4
	3	-5.7	-4.0	43.7
	4	-8.8	-3.8	20.0
	5	-4.4	**-0.8**	**19.2**

Table 2: Relative bias in per cent for different combinations of variance estimators and x-vectors, for study variables y_1-y_3. The lowest absolute relative bias

for each study variable and variance estimator is highlighted.

In Table 3, the actual coverage rates of 95 per cent confidence intervals for μ, based on the propensity score estimator and the different variance estimators, are presented. We see that in almost all cases, the actual coverage rate is far below 95 per cent. This is especially true when **x**-vector M4 or M5 is used. Throughout, the coverage rates for the JK method are the largest, and the closes to the desired coverage rate. From Table 2, the JK method overestimates the true variance. In effect, the JK confidence intervals are often wide enough to embrace the true value even though the point estimator is biased.

Study variable	v	POST	a MOD	JK
y_1	1	90.9	91.3	96.1
	2	83.5	84.0	95.2
	3	60.0	60.1	81.4
	4	44.4	45.5	59.6
	5	59.4	60.0	67.5
y_2	1	65.6	67.3	87.7
	2	68.7	69.3	89.6
	3	69.7	70.1	86.9
	4	13.7	14.2	25.0
	5	13.1	13.7	20.8
y_3	1	50.2	52.0	80.0
	2	70.2	70.8	89.7
	3	67.5	67.6	86.5
	4	64.1	64.7	75.4
	5	25.6	26.1	33.5

Table 3: Coverage rate in per cent for various combinations of estimation approach and **x**-vector, for study variables y_1-y_3.

5. Conclusions and Ideas for Future Work

Our simulation results suggest, not very surprisingly, that the propensity score estimator represents a less biased alternative to the Web sample mean as estimator of the population mean. The choice of **x**-vector seems to really matter: in our study, the **x**-vectors M1 and M2 reduced the bias of the propensity score point estimator, whereas M4 and M5 reduced the biases of the suggested variance estimators. The use of the last-mentioned **x**-vectors, however, also gave far too low coverage rates. The JK variance estimator stood out by severely overestimating the true variance, but also by producing coverage rates that were closest to the desired 95 per cent level.

Some possible expansions of our study are the inclusion of continuous study variables and the use of bootstrap for variance estimation.

6. Acknowledgment

7. References

D'Agostino, R.B., Jr and Rubin, D.B. (2000), Estimating and using propensity scores with partially missing data, *Journal of the American Statistical Association*, 95, 749-759.

Beissel-Durrant, G. and Skinner, C. (2004), Using missing data methods to correct for measurement error: An application to the distribution of hourly pay, paper presented at the *Monday Afternoon Seminar Nov 1, 2004, Institute for Social & Economic Research*, University of Essex, UK.

Cassel, C. M., Särndal, C.-E. and Wretman, J. H. (1983), Some uses of statistical models in connection with the nonresponse problem, in *Incomplete Data in Sample Surveys, Vol II,* Academic Press: New York, 143-160.

Cochran, W. G. (1968), The effectiveness of adjustment by subclassification in removing bias in observational studies, *Biometrics*, 24, 295-313.

Cook, M. L. and Rafferty, A. P. (2004), *Health Risk Behaviors in the State of Michigan: 2003 Behavioral Risk Factor Survey*. Lansing, MI: Michigan Department of Community Health, Bureau of Epidemiology, Epidemiology Services Division.

Isaksson, A., Danielsson, S. and Forsman, G. (2004), On the variability of estimates based on propensity score weighted data from Web panels, *2004 Proceedings of the American Statistical Association*, Section on Survey Research Methods [CD-ROM], Alexandria, VA: American Statistical Association, 3689-3696.

Lee, S. (2004), *Statistical Estimation Methods in Volunteer Panel Web Surveys*. Doctoral dissertation, Joint Program in Survey Methodology, University of Maryland, USA.

Little, R. J. A. (1986), Survey nonresponse adjustments, *International Statistical Review*, 54, 139-157.

Little, R. J. A. and Rubin, D. B. (1987), *Statistical Analysis with Missing Data*. New York: Wiley.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996) *Applied Linear Statistical Models*, 4th edn. Chicago: Irwin.

Oh, H. L. and Scheuren, F. S. (1983), Weighting adjustments for unit nonresponse, in *Incomplete Data in Sample Surveys, Vol II: Theory and Anno-*

tated Bibliography (W. G. Madow, I. Olkin and D. B. Rubin, eds). New York: Academic Press.

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

Rosenbaum, P.R. and Rubin, D.B. (1984), Reducing bias in observational studies using subclassification on the propensity score, *Journal of the American Statistical Association*, 79, 516-524.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. New York: Springer.

Terhanian, G., Smith, R., Bremer, J., and Thomas, R. K. (2001), Exploiting analytical advances: Minimizing the biases associated with Internet-based surveys of non-random samples, *ARF/ESOMAR: Worldwide Online Measurement*, ESOMAR Publication Services, 248, 247-272.

Wolter, K. M. (1985), *Introduction to Variance Estimation*. New York: Springer.

Appendix: Derivations of Poststratification Formulas

We start with eqn. (3), which is derived from Särndal et al. (1992, eqn. (7.6.1)):

$$\hat{\bar{y}}_U = \frac{\hat{t}_{yr}}{N} = \sum_{h=1}^{H} \frac{N_h}{N} \widetilde{y}_{s_{Wh}} = \sum_{h=1}^{H} \frac{N_h}{N} \frac{\sum_{s_{Wh}} \frac{y_k}{\pi_k^*}}{\sum_{s_{Wh}} \frac{1}{\pi_k^*}} = \sum_{h=1}^{H} \frac{N_h}{N} \frac{\frac{Na_h}{n_W}\sum_{s_{Wh}} y_k}{\frac{Na_h}{n_W} n_{Wh}} = \sum_{h=1}^{H} \frac{N_h}{N} \bar{y}_{s_{Wh}}$$

We now turn to eqn. (4), which is derived from Särndal et al. (1992, eqn. (7.2.11)) by setting, for $h=1,\ldots,H$,

$$e_{ks} = y_k - \bar{y}_{s_{Wh}}; k \in s_{Wh}$$

$$g_{ks} = \frac{N_h}{n_{Wh}} \pi_h^*; k \in s_{Wh}$$

$$\breve{\Delta}_{kl} = \frac{n - Na_h}{Na_h(n-1)}; k \neq l \in s_{W_h}$$

$$\breve{\Delta}_{kl} = 1 - \frac{n}{Na_h}; k = l \in s_{W_h}$$

We then have:

$$\hat{V}(\hat{\bar{y}}_U) = \frac{1}{N^2} \sum_{h=1}^{H} \sum\sum\sum_{s_{Wh}} \breve{\Delta}_{kl} \left(g_{ks} \frac{e_{ks}}{\pi_h^*} \right) \left(g_{ls} \frac{e_{ls}}{\pi_h^*} \right)$$

$$= \frac{1}{N^2} \sum_{h=1}^{H} (1 - \pi_h^*) \frac{N_h}{n_{Wh}} \sum_{s_{Wh}} (y_k - \bar{y}_{s_{Wh}})^2$$

$$- \frac{1}{N^2} \sum_{h=1}^{H} \frac{n - Na_h}{Na_h(n-1)} \frac{N_h}{n_{Wh}} \sum_{s_{Wh}} (y_k - \bar{y}_{s_{Wh}})^2$$

$$= \sum_{h=1}^{H} (1 - \pi_h^*) \left(\frac{N_h}{N} \right)^2 \frac{s_{Wh}^2}{n_{Wh}} \left[\frac{n_h - 1}{n_h} \left(1 - \frac{n - Na_h}{Na_h(n-1)} \frac{1}{1 - \pi_h^*} \right) \right]$$

$$\approx \sum_{h=1}^{H} \left(1 - \frac{n_W}{Na_h} \right) \left(\frac{N_h}{N} \right)^2 \frac{s_{Wh}^2}{n_{Wh}}$$

where the last approximate equality holds since

$$\frac{n_h - 1}{n_h} \left(1 - \frac{n - Na_h}{Na_h(n-1)} \frac{1}{1 - \pi_h^*} \right)$$

$$= \frac{n_h - 1}{n_h} \left(1 - \frac{n - Na_h}{Na_h(n-1)} \frac{Na_h}{Na_h - n} \right)$$

$$= \frac{n_h - 1}{n_h} \frac{n}{n-1} \approx 1$$