



## Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys

F. Kreuter,

*University of Maryland, College Park, USA*

K. Olson,

*University of Nebraska, Lincoln, USA*

J. Wagner,

*University of Michigan, Ann Arbor, USA*

T. Yan,

*National Opinion Research Center, Chicago, USA*

T. M. Ezzati-Rice,

*Agency for Healthcare Research and Quality, Rockville, USA*

C. Casas-Cordero and M. Lemay,

*University of Maryland, College Park, USA*

A. Peytchev

*RTI International, Research Triangle Park, USA*

and R. M. Groves and T. E. Raghunathan

*University of Michigan, Ann Arbor, USA*

[Received August 2008. Final revision July 2009]

**Summary.** Non-response weighting is a commonly used method to adjust for bias due to unit non-response in surveys. Theory and simulations show that, to reduce bias effectively without increasing variance, a covariate that is used for non-response weighting adjustment needs to be highly associated with both the response indicator and the survey outcome variable. In practice, these requirements pose a challenge that is often overlooked, because those covariates are often not observed or may not exist. Surveys have recently begun to collect supplementary data, such as interviewer observations and other proxy measures of key survey outcome variables. To the extent that these auxiliary variables are highly correlated with the actual outcomes, these variables are promising candidates for non-response adjustment. In the present study, we examine traditional covariates and new auxiliary variables for the National Survey of Family

*Address for correspondence:* F. Kreuter, Joint Program in Survey Methodology, 1218 LeFrak Hall, University of Maryland, College Park, MD 20742, USA.  
E-mail: fkreuter@survey.umd.edu

Growth, the Medical Expenditure Panel Survey, the American National Election Survey, the European Social Surveys and the University of Michigan Transportation Research Institute survey. We provide empirical estimates of the association between proxy measures and response to the survey request as well as the actual survey outcome variables. We also compare unweighted and weighted estimates under various non-response models. Our results from multiple surveys with multiple recruitment protocols from multiple organizations on multiple topics show the difficulty of finding suitable covariates for non-response adjustment and the need to improve the quality of auxiliary data.

*Keywords:* Interviewer observations; Non-response adjustment; Non-response bias; Paradata; Response propensity weights

## 1. Introduction

Household surveys in many countries have witnessed a decline in response rates over the past few decades (Atrostic *et al.*, 2001; Curtin *et al.*, 2005; De Leeuw and DeHeer, 2002). Declining response rates raise concerns among survey organizations and data users about the non-response bias and precision of survey estimates. The danger of a low response rate is the presence of non-response bias if sampled people who are unlikely to participate in a survey differ *systematically* from likely participants *with regard to survey outcome variables of interest*. Survey organizations therefore aim to do their best to minimize potential non-response bias that is associated with survey statistics.

One strategy for reducing non-response bias involves investing extensive resources in contacting sample subjects and persuading them to participate in the survey. Techniques such as advance letters, incentives and customized call scheduling and calling rules have been shown empirically to be effective in increasing response rates at the data collection stage, although their use has not been effective at stopping the decline in response rates in the past 15 years (Dillman *et al.*, 2002). The extensive effort spent on contacting and recruiting sample subjects with low probability of responding is costly. Given standard resources and budgets, survey organizations are limited in the extent to which they can pursue those with low response probability (Rogers *et al.*, 2004; Groves and Peytcheva, 2008).

After data have been collected a strategy to address potential non-response bias is the construction of post-survey adjustments. Weighting is one such post-survey adjustment method. With non-response weighting, whether it is a weighting class adjustment method (Little, 1986; Kalton and Flores-Cervantes, 2003) or a response propensity weighting method (Rosenbaum and Rubin, 1983), survey respondents are assigned a weight to compensate for their differential probability of participation given selection into the sample. These weighting adjustments are aimed at reducing non-response bias in the final survey estimates, although they often increase the variance of the adjusted estimate as well.

The success of non-response weighting depends on the variables that are used in constructing non-response weights. Adjustment variables that are most successful at reducing non-response bias have two properties (Little, 1986; Kalton and Flores-Cervantes, 2003; Little and Vartivarian, 2003, 2005; Groves, 2006). The variables should be predictive

- (a) of the sampled person's probability of responding to a survey request and
- (b) of the survey outcome variables of interest.

The latter criterion has sometimes been neglected in the past, but some common cause of both the survey outcome variables and survey participation is necessary for non-response bias to appear (Groves, 2006). A recent simulation study by Little and Vartivarian (2005) demonstrated that weighting was effective in substantially reducing non-response bias without increasing the estimated variance only when the variables that are used in constructing weights were highly

correlated with both the survey variables of interest and the response propensity. Whether such correlations can be found in practice is one of the research questions that we address in this paper.

Only variables that are available for both responding and non-responding cases can be used in non-response weighting adjustments. Often very few such variables are available. In area probability surveys, geographic variables such as region, primary sampling unit and neighbourhood are available. In telephone surveys, area codes and exchanges are available. Some researchers have explored the use of variables that are available at the cluster level, such as average income or percentage of minority residents in a given geographical unit (Kalton and Flores-Cervantes, 2003; Johnson *et al.*, 2006). However, much of this research has not explored the strength of the association between these or other demographic variables that are associated with survey outcomes and the probability of responding to the survey request. When auxiliary variables are only weakly associated with the survey variables of interest and the response to the survey request, using them in a weighting adjustment may not reduce non-response bias and may increase variance. Although many response propensity models that are used in practice show low predictive power as evidence by goodness-of-fit measures such as pseudo- $R^2$ , they can help to adjust for non-response bias when the predictors of the response probability are associated with the survey outcome variables.

Promising *alternative* sets of auxiliary variables are also available but so far they have been only rarely considered for weighting purposes. These auxiliary variables include those found in augmented sampling frames or previous surveys of the same sample, as well as those collected for survey management purposes or collected by interviewers at little extra expense. The actual correlation between these alternative weighting adjustment variables with survey outcomes is an open empirical question, as is their correlation with survey participation.

For government agencies or other survey organizations that are interested in a set of uniform adjustment procedures across multiple surveys, such as in multinational research, data that are collected during the field period may pose a unique solution. A typical example would be the European Social Survey (ESS) (<http://www.europeansocialsurvey.org/>). Here, the same survey is fielded in various European countries, but sampling frame information that is available for non-response adjustment varies substantially across countries. It is likely that for international comparative research comparable data collection procedures are easier to implement than the construction of comparable sampling frames and sampling frame information.

To summarize, this paper has three aims, each addressed in one of the following sections.

- (a) Using five large-scale surveys as examples, we shall introduce alternative sources for non-response adjustment variables in Section 2. For each of the five surveys, we shall describe the key survey outcome variables  $y$  and identify auxiliary variables  $z$  that have the potential to be used for non-response adjustment. We shall give examples for data collected during the recruitment process as a potential source for non-response adjustment variables. Quality of measurement of those alternative variables will be discussed at the end in the context of our findings.
- (b) In Section 3, we shall examine the correlation of each of these auxiliary variables  $z$  with the response indicator  $r$  and survey outcome variables of interest ( $y$ ). The auxiliary variables are available for both respondents ( $r = 1$ ) and non-respondents ( $r = 0$ ). Thus correlations between  $z$  and  $r$  can be examined for the entire sample. The correlations between auxiliary variables  $z$  and survey outcomes  $y$ , in contrast, can only be observed for respondent cases ( $r = 1$ ). To the extent that the auxiliary  $z$ -variables are correlated with both the response indicator  $r$  and the survey outcome variable  $y$ , they can be useful adjustment variables.

- (c) Having examined the auxiliary variables  $z$  individually, in Section 4 we then ask the following questions: when  $z$ s are added to the covariates  $x$  that are traditionally used for non-response adjustment, does the resulting new set of weights change the estimated non-response-adjusted means of the key survey outcome variables  $y$ ? And, if so, does the change reflect a reduction in non-response bias, and what are the consequences for the estimated variances?

The effectiveness of sample-based weighting non-response adjustments depends on whether the respondents and non-respondents are comparable conditional on the information that is used in the non-response adjustment, i.e. that the outcome variables  $y$  are missing at random for non-respondents. In most practical situations, this is unknowable as information on the outcome variables of interest  $y$  is not available for both respondents and non-respondents. Here, we examine auxiliary variables which we assume increase the chance of meeting the missingness at random missing data mechanism assumption for some of the outcome variables. If the missing data mechanism is data not missing at random, then other adjustment methods are needed to remove non-response bias in the adjusted mean (Little and Rubin, 2002).

## 2. Five projects with alternative adjustment variables

All five projects focus on non-response adjustment using variables that are available for every sample subject. In an attempt to find good auxiliary variables for non-response adjustment (ideally proxy measures of the survey outcome variables), each of these five projects turned to data from one of two sources: either *enriched sampling frame* information available from external administrative records or previous surveys of the same sample units or *paradata* such as interviewer observations and call record data.

### 2.1. Enriched sampling frame information

The first two projects use data from enriched sampling frames to test alternative non-response adjustment models. Sampling frames usually provide only a limited number of variables for weighting adjustment. In area probability samples, information on all sample units is often available only for a census tract or higher geographic levels (Groves and Couper, 1998). However, if the sampling frame is a population register or administrative list of some sort, sampling frames can also contain rich individual level data that are related to the survey outcome or are even proxy measures of survey variables. Examples of such sampling frames are medical provider records or voter registries. Increasingly, sampling frames are also enriched by linking sampled cases to administrative data that are related to the survey outcome variables. Other situations in which sampling frames provide rich data are samples drawn from respondents to previous surveys. If the topic of the subsequent survey is similar, these sources of data provide auxiliary variables ( $z$ -variables in this context) that are likely to be related to the key survey outcome variables. The usefulness of these  $z$ -variables for non-response adjustment will be examined here. A summary of the  $z$ -variables is given in Table 1.

- (a) The first project uses data from a survey that was conducted by the University of Michigan Transportation Research Institute (UMTRI) (Bingham, 2007). This is a survey of young adults who completed high school and have a valid Michigan driver's licence. A large sample of young adults was selected and interviewed while still in high school. A year later, they were contacted for a follow-up interview by telephone. The UMTRI was able to enrich the sampling frame of this survey with driving record data from the State of

**Table 1.** Studies and new adjustment variables

<i>Study</i>	<i>z-variable(s)</i>	<i>Description</i>
University of Michigan Transportation Research Institute	Traffic points in past 12 months	Count
	Serious traffic offences in past 12 months	Count
	Traffic offences in past 12 months	Count
	Crashes in past 12 months	Count
	Serious crashes in past 12 months	Count
Medical Expenditure Panel Survey	Dwelling unit member not working for health reasons	Binary
European Social Survey	Nights of dwelling unit members in the hospital	Count
	Litter observed around sampled address	Binary
American National Election Survey	Multiunit housing structure	Binary
	Negative comment during recruitment	Binary
National Survey of Family Growth	Single-family home	Binary
	Guess on respondent's sexual activity	Binary
	Evidence of children in the household	Binary

Michigan for all sampled young adults. Data from the telephone interview are available for about 40% of all young adults for whom data from the Michigan Department of Motor Vehicles are available ( $n = 12\,694$ ). The survey asks the sampled young adults about various driving and risk-taking behaviours. Information on age and sex of the sample cases is available from the sampling frame and is currently used for non-response adjustment. The enriched sampling frame contains five official record variables that fulfil the criteria for the alternative non-response weighting models that are examined in this paper. The  $z$ -variables of 'traffic points received in the past 12 months, number of traffic offences, number of serious traffic offences, number of crashes and number of serious crashes in the past 12 months' are likely to be related to key survey outcome variables such as seat belt wearing, traffic violation and revoked licences.

- (b) The second project uses data from the Medical Expenditure Panel Survey (MEPS), which is sponsored by the Agency for Healthcare Research and Quality. The MEPS is conducted to produce national estimates of health care use, expenditures, sources of payment and insurance coverage for the civilian non-institutionalized US population. Each annual sample of households for the MEPS is drawn from respondents to the prior year's National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics of the Centers for Disease Control and Prevention (Cohen, 1997; Ezzati-Rice *et al.*, 2008). The household component of the MEPS provides annual data constructed from two consecutive overlapping panels with approximately 13000 dwelling units and 35000 people. The MEPS dwelling unit response rate averages about 80% (among the NHIS households that were fielded for the MEPS). In the MEPS, the first level of adjustment is at the level of the dwelling unit, which is used to adjust for non-response among those households that were subsampled from the NHIS for the MEPS. The variables that were historically used for non-response adjustment include demographic, socio-economic and geographic variables as collected during the prior year's NHIS interview (in detail: race/ethnicity, marital status, gender and education of the household reference person, number of people in the dwelling unit, family income, census region, area size, metropolitan and urban-rural indicators, type of primary sampling unit, sampling domains of predicted poverty status and any black in the household, interruption in telephone service

as well as home ownership status). Starting with the MEPS panel that was introduced in 2004, health-related variables as collected in the prior year's NHIS have also been included in the MEPS dwelling unit non-response adjustment. In the alternative non-response adjustment models that are considered below, the two health-related  $z$ -variables of 'not working for health reasons' and 'number of nights in the hospital' are used in addition to the standard set of variables evaluated for non-response adjustment (Wun and Ezzati-Rice, 2007). The non-response-adjusted MEPS weights can be used to examine how well their use reproduces selected NHIS fully weighted variables that include both MEPS respondents and non-respondents. These NHIS estimates are highly correlated with key outcome variables in the MEPS. In particular, three key NHIS survey outcome variables are of interest here: proportion of responding dwelling units with no member with doctor's visits in the past 2 weeks, no dwelling unit member with a limitation in daily activities and no dwelling unit member with a barrier to healthcare due to cost in the past 12 months. In addition, a constructed dollar-denominated index for the amount of expected expenditures derived from qualitative health status will also be used in the evaluation.

## 2.2. *Paradata from field efforts*

Largely untapped sources of information that could be used for non-response adjustment are paradata (Couper, 1998), such as interviewer observations and contact record data, collected during fieldwork. Adding these variables to the adjustment procedures is possible with little expense to researchers and might help to reduce non-response bias. Data collection organizations only recently started to collect paradata. The US Census Bureau now uses an automated system for collecting contact histories for their computer-assisted personal interviews (Bates *et al.*, 2008). Other government statistical agencies have started to use similar procedures. These data are gathered primarily to provide progress feedback to field operations (Groves and Couper, 1998; Groves and McGonagle, 2001). Several in-person surveys collect even more information through interviewer observation at the neighbourhood, housing unit and contact level.

There are several examples of studies in which these paradata have been used to estimate survey participation. Observations about the type of housing (e.g. multiunit structures) and the presence of door intercoms have shown to be predictive of the amount of effort that is required to contact sample households (Groves and Couper, 1998; Lynn, 2003). Contact information from the 2005 NHIS in the USA was used to predict participation in the NHIS (Bates *et al.*, 2008). Likewise, interviewer-assessed enthusiasm of sample members about participating in the British National Survey of Sexual Attitudes and Lifestyles was successful in predicting their probability of response (Copas and Farewell, 1998). However, as discussed earlier, the correlation with the probability of responding that was demonstrated in the above studies is only one aspect of good adjustment variables. Only when these paradata are also related to key survey outcome variables will they be useful for non-response bias adjustment. Whether or not paradata can be useful auxiliary  $z$ -variables in this context—i.e. whether paradata variables are also related to survey variables—is examined in three of the projects that were used in the current study.

- (a) The ESS is a multipurpose face-to-face survey that takes place every 2 years. It is designed to measure and monitor changing social attitudes and values across Europe, which are then used to identify and interpret trends over time. The 2002 ESS outcome variables span constructs such as social trust, politics and citizen involvement. About half of the countries participating in the first round of the ESS in 2002 came close to the specified target response rate of 70%. Greece, Poland and Portugal, the three countries

that were used for this project, have response rates of 80%, 72% and 69% respectively. Currently, the ESS performs no routine non-response adjustment. All interviewers were instructed to collect information on the type of housing structure in which the sample units lived and whether an alarm system, intercom, entry phone, security lights, porch and so on were visible on the housing unit. Such impediments to access are potentially related to non-response (Groves and Couper, 1998). The interviewer also recorded how common litter and garbage were in the area of the sample address. These neighbourhood observations are seen as correlates of survey items about social involvement and more general social trust (Schnell, 1997; Couper *et al.*, 1998; Groves and Couper, 1998; Abraham *et al.*, 2006). For more information on the ESS see <http://www.europeansocialsurvey.org/>.

- (b) The 2004 American National Election Study (ANES) (University of Michigan, 2004) is another example of a survey with paradata that are collected by the interviewer. The 2004 ANES is a face-to-face survey of the adult US population. In this investigation, we use only data that were collected before the Bush–Kerry presidential election. In the pre-election interview, 1212 sample subjects were respondents and 621 were non-respondents. The traditional ANES non-response adjustment is a weighting class adjustment using US census regions and a metropolitan statistical area indicator. A unique feature of the ANES is the collection of interviewer observations on respondents and non-respondents. The interviewer observations document what the informant or sample unit said ‘on the doorstep’ to the interviewer. Two of these interviewer observations are employed in the alternative non-response adjustment model that is examined here. For all households contacted the interviewer recorded whether the household informant made a negative comment about the survey request. Examples for such negative comments are expressions such as ‘surveys are a waste of time’, ‘I don’t trust surveys’, ‘surveys are a waste of taxpayers money’ or ‘I’m not interested’. This indicator of negative comments is a  $z$ -variable that has been shown to be related to both participation in the survey and the key survey outcome variables on interest in politics and voting (Couper, 1997). The type of housing structure (single-family home, multiunit housing, etc.) also was observed for all sample households and will be used in this analysis.
- (c) The National Survey of Family Growth (NSFG) (cycle 7) is our third example of a survey using paradata that are collected by the interviewer. The survey is based on a sample of the household population of the USA, 15–44 years of age. The NSFG is conducted as a continuous survey. Fieldwork is carried out by the University of Michigan’s Institute for Social Research under a contract with the National Center for Health Statistics. Three-quarters of an annual sample are used here; the weighted response rate is about 75%. The current NSFG non-response adjustment includes observations that were obtained during the housing unit listing process and interviewer observations from all contacts to the household (Lepkowski *et al.*, 2006). Recently NSFG interviewers were asked to record an additional  $z$ -variable that should be strongly related to the survey outcome of interest. The NSFG interviewers assess whether each sample person contacted (both respondents and refusals) is sexually active or not, i.e. interviewers are asked to record their answer to the following question: ‘Do you think the selected respondent is in an active sexual relationship with an opposite-sex partner?’. In the NSFG training interviewers are made aware that being in an active sexual relationship with an opposite sex partner is a strong predictor of important variables in the NSFG (such as pregnancy) and are told that their observations will be used for non-response analysis in case a completed main interview with the selected respondent cannot be secured. The interviewer observation can be seen

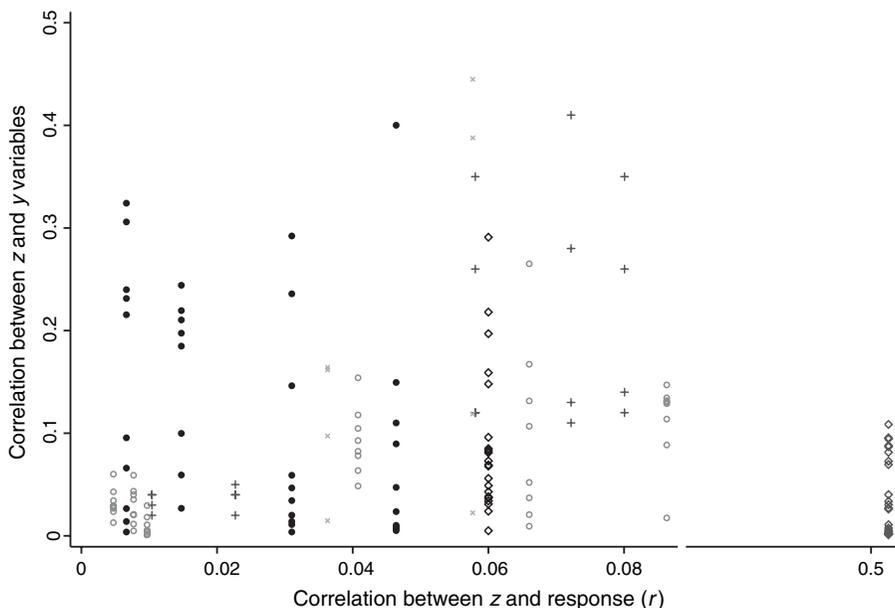
as a proxy measure of key NSFG variables and will be examined here in addition to existing adjustment variables. In addition, the interviewer recorded whether there was any evidence of children in the household. This also is a potential proxy indicator for key survey outcome variables and will be used as a  $z$ -variable in this paper.

### 3. Correlations of auxiliary variables with response indicator and survey variables

Fig. 1 displays the bivariate correlations estimated for each of the five projects. The  $x$ -axis in Fig. 1 reflects the correlation between the response indicator  $r$  and the additional adjustment variables  $z$  for each of the surveys. Surveys in which multiple  $z$ -variables are available appear multiple times along the  $x$ -axis. The  $y$ -axis presents the correlation between the respective  $z$ -variable and a selected set of survey variables  $y$ , which are listed in Tables 2–6.

All sample cases for which auxiliary variables from either frame information or paradata are available are used to estimate the correlations between the auxiliary variables and the response indicator  $\rho(z, r)$ . To the extent that the auxiliary variables are available only for people, contacted already as for the interviewer observation in the NSFG and the ANES measure, the correlation  $\rho(z, r)$  is estimated only for the subset of contacted people. For the MEPS  $\rho(z, y)$  can be estimated for the full sample because  $y$ -variables are part of the NHIS and therefore available for all MEPS sample cases whether responding or not. For all other surveys the correlations with the survey outcome ( $\rho(z, y)$ ) are estimated for respondents only.

A few observations can be made about Fig. 1. First, most of the auxiliary  $z$ -variables are only very weakly correlated with the response indicator. Only for the ANES is the estimate for  $|\rho(z, r)|$  greater than 0.1. The ANES  $z$ -variable ‘negative comment made by household’ has a correlation of about 0.51 with the response indicator. Second, for any given estimate of  $|\rho(z, r)|$ ,  $|\rho(z, y)|$  varies remarkably across the different  $y$ -variables. This indicates the difficulty in finding a single suitable adjustment variable with the potential to reduce non-response bias on several survey



**Fig. 1.** Relationship between the correlation of  $z$ - and  $y$ -variables and correlation of  $z$  and response in five surveys (all correlations are shown as absolute values): +, UMTRI; x, MEPS; O, ESS;  $\diamond$ , ANES;  $\bullet$ , NSFG

**Table 2.** Survey outcome variables from the UMTRI survey

<i>y</i> -variable	Description
Number of traffic tickets	Count
Licence ever revoked	Binary
Never wearing seat belt on local travels	Binary
Never wearing seat belt on long travels	Binary

**Table 3.** Survey outcome variables from the MEPS

<i>y</i> -variable	Description
Dollar index derived from health status	Continuous
Doctor visits in past 12 months	Binary
Limitations in daily activities	Binary
Barriers to healthcare due to cost	Binary

outcomes. Third, most of the estimated  $|\rho(z, y)|$  correlations are smaller than 0.2. In the ESS the interviewer observation of litter in the sample unit's neighbourhood is only weakly correlated with any of the outcome measures ( $|\rho(z, y)| < 0.1$ ) for all three countries (Greece, Poland and Portugal). For the UMTRI, MEPS and NSFG stronger correlations are observed for some of the *y*-variables. However, these correlations are not as high as the Little and Vartivarian (2005) simulations find are necessary to lead to substantial reduction in both bias and variance of an adjusted respondent mean. Fourth, for the MEPS the estimated correlations  $\rho(z, y)$  that are displayed in Fig. 1 reflect not just the correlation between the respondent cases but for all sample elements. It is noteworthy that the range and size of relationships that are observed here are similar to those found in the four other projects (the NSFG, ANES, UMTRI and ESS) where information is available only for responding sample units.

#### 4. Effects on adjusted estimates

So far the auxiliary variables show only modest potential to be useful for non-response adjustment. What matters for practitioners is whether adding these auxiliary variables to the existing weighting procedures will lead to changes in the estimated means and subsequently whether those changes translate into a substantial reduction in non-response bias without increasing the variance.

##### 4.1. Changes in estimated means

Next, we compare estimated non-response-adjusted means for weights with the new adjustment variables ( $w_{xz}$ ) and without ( $w_x$ ) these new variables. We then relate the change in the weighted estimates to the previously discussed correlations. For all surveys that deviate from an equal probability sample design, we also use selection weights. We refer to the respondent means as  $\bar{y}$  if they are unadjusted for non-response.

Different weighting methods are used in the five surveys to match most closely the weighting procedure that is usually performed in each of the surveys. For the two projects in which rich

**Table 4.** Survey outcome variables from the ESS

<i>y</i> -variable	Description
Most people can be trusted	0–10 scale
Most people try to be fair	0–10 scale
Most of the time people are helpful	0–10 scale
Trust in country's Parliament; legal system	0–10 scale
Trust in the European Parliament; United Nations	0–10 scale
Interested in politics	Binary
Politics are not complicated to understand	Binary
Could take an active role in a group involved with political issues	Binary
Easy to make mind up about political issues	Binary
Politicians care about what people like you think	Binary
Politicians are interested in people's opinions	Binary
Not victim of burglary or assault in last 5 years	Binary
Feeling of safety when walking alone in local area after dark	Binary
Good subjective general health	Binary
Not hampered in daily activities	Binary

**Table 5.** Survey outcome variables for the ANES

<i>y</i> -variable	Description
Interested in campaigns	5-point scale
Respondent voted in 2000	Binary
Number of days watched national news on television in the last week	Count
Respondent cares who wins House election	Binary
General attitude towards John Edwards (candidate in the election)	0–100 scale
Is there anything respondent likes about the Republican Party	Binary
Respondent put off medical treatment owing to cost	Binary
Affect for G. W. Bush: G. W. Bush made respondent feel hopeful	Binary
Respondents' self-placement on liberal–conservative dimension	5-point scale
Care who wins Presidential election	Binary
Unemployment better or worse in last year	5-point scale
Defence spending scale: Democratic Party placement	7-point scale
Importance of government health insurance issue to respondent	5-point scale
Attitude towards abortions	Binary
Respondent intention to vote this November (2004)	Binary
Religion important part of respondent's life	Binary
Active at church besides attendance	Binary
Respondent's age	Continuous
Both parents born in USA	Binary
Years lived in current home	Count

frame data are available (the MEPS and UMTRI), the non-response weight is formed out of the predicted response probabilities  $\hat{p}$  from a logistic regression model, where  $\hat{p} = \exp(\lambda) / \{1 + \exp(\lambda)\}$ . For the MEPS and UMTRI projects the logit is estimated either by using the original set of adjustment covariates (thus  $\lambda = \hat{\lambda}_x = X'_i \hat{\beta}_x$ ) or a combined set of covariates which includes the original and the alternative adjustment variables, i.e.  $\hat{\lambda} = \hat{\lambda}_{xz} = X'_i \hat{\beta}_x + Z'_i \hat{\beta}_z$  (Table 7). Adjustment cells are formed by using means within quintiles of the estimated response probabilities (Little, 1986; Little and Rubin, 2002), and the inverse of the average predicted probabilities within each quintile is used as a weight (Wun and Ezzati-Rice, 2007; Yan and Raghunathan, 2007).

**Table 6.** Survey outcome variables for the NSFG†

<i>y</i> -variables	Description
Age of respondent	Continuous
Total number of live births given by female respondent (♀)	Count
Additional births expected in the future (♀)	Binary
Respondent's age at last or most recent sexual intercourse (♀)	Continuous
Total number of pregnancies (♀)	Count
Number of opposite sex partners in lifetime (♀)	Count
Age of respondent's first sexual partner at time of respondent's first sex (♀)	Continuous
Intercourse in 3 months prior to interview (♀)	Binary
Number of children in household (♂)	Count
Number of completed pregnancies respondent has fathered (♂)	Count
Number of spontaneous pregnancy losses fathered by respondent (♂)	Count
Number of abortions fathered by respondent (♂)	Count
Ever married or cohabited (♂)	Binary
Ever had intercourse (♂)	Binary
Ever cohabited (♂)	Count
0–99% of poverty level (♂)	Continuous
2 biological or adopted parents from birth (♂)	Binary

†Different items are available for males and females, which are indicated with the gender symbol in the table.

In the NSFG, similar logistic regression models are currently used to create non-response adjustment weights (Lepkowski *et al.*, 2006). Response propensities are estimated for the probability to respond to a screening interview used to identify eligible people within sampled households. The inverse of this estimated propensity is used as a weight. The weights in the lowest decile are trimmed to the median value for that decile (Little, 1986). To create the new non-response weight, a weighting class adjustment is also implemented. Conditional on having completed a screening interview, cases are assigned to two classes by using the binary  $z$ -variable (Groves *et al.*, 2007). The inverse of the response rate within each cell is used as an adjustment factor. This adjustment factor is then multiplied by the inverse of the trimmed estimated screening interview response propensity and the probability-of-selection weight.

For the ESS and ANES, weighting class adjustments are employed (Kreuter *et al.*, 2007; Peytchev and Olson, 2007). For the ESS no non-response adjustment is currently used; thus we use each of the dichotomous  $z$ -variables separately to form alternative two non-response adjustment classes for each variable. In the ANES, the dichotomous  $z$ -variables were added separately to the non-response adjustment, crossing the existing adjustment cells ( $c = 8$  classes that are formed out of the two  $x$ -covariates) by the additional auxiliary variable  $z$ . The number of respondents in class  $c$  is indicated by  $n_{rc}$  in Table 7. All sampled elements in that class are indicated by  $n_c$ . The inverse of the estimated response rate  $\phi_c = n_{rc}/n_c$  in each of the cells is the existing non-response weight. The inverse of the estimated response probability  $\phi_{cz}$ —the ratio of respondents in the cell that is formed by including  $z$  ( $n_{rcz}$ ) to all sample elements in the cell that is formed by including  $z$  ( $n_{cz}$ )—is the new non-response weight.

To compare the effect of incorporating the  $z$ -variables into the weights on various estimates, we express the change in the adjustment mean in standard error units:

$$\left| \frac{\hat{y}_{w_{xz}} - \hat{y}_{w_x}}{\text{SE}(\hat{y}_{w_x})} \right| \quad (1)$$

**Table 7.** Non-response adjustment weights

<i>Adjustment weight</i>	<i>Logit in propensity model</i>	<i>Weighting class adjustment</i>
$w_x$	$\hat{\lambda}_x = X'_i \hat{\beta}_x$	$\phi_c = n_{rc} / n_c$
$w_{xz}$	$\hat{\lambda}_{xz} = X'_i \hat{\beta}_x + Z'_i \hat{\beta}_z$	$\phi_{cz} = n_{rcz} / n_{cz}$

where  $\hat{y}_{w_{xz}}$  is the estimated mean weighted by the selection weight and the new non-response adjustment (including both old covariates  $x$  and new auxiliary variables  $z$ ),  $\hat{y}_{w_x}$  is the estimated mean weighted with the traditional non-response adjustment variables and the selection weight, and  $SE(\hat{y}_{w_x})$  is the estimated design-based standard error for the estimate of  $y$  by using the traditional non-response adjustment. The standardized difference in the weighted estimate allows us to put all the estimates on the same scale. For the ESS this is just the difference between the non-response-adjusted estimated mean and the unadjusted estimated respondent mean, since no prior adjustment scheme exists.

Fig. 2 shows the effect of using interviewer observations in the non-response adjustment for the ESS. The  $x$ -axis represents the correlation between the respective interviewer observations and the survey variables, which for the ESS is simply the values that are plotted on the  $y$ -axis from Fig. 1. The  $y$ -axis of this graph displays the difference of the weighted means with and without non-response adjustment in standard error units (equation (1)). As expected from theory and prior simulations the change in the weighted estimates increases with larger correlations between  $z$  and  $y$ . To give one example, for Poland the correlation between the interviewer observation of the respondent’s ‘housing structure’ and the respondent’s ‘feeling of safety when walking alone after dark in the local area’ is  $|\rho(z, y)| = 0.265$  and led to a shift in the weighted mean of about 1.4 standard error units of the  $y$ -variable.

For all other surveys prior adjustment variables exist and thus it would be inappropriate to use the same  $x$ -axis as we did for the ESS. Instead we plot the standardized difference from equation (1) against the difference in the correlation between the predicted response probabilities (with and without the alternative auxiliary adjustment variables  $z$ ) and the survey outcome variables. In the presence of multiple adjustment variables we use

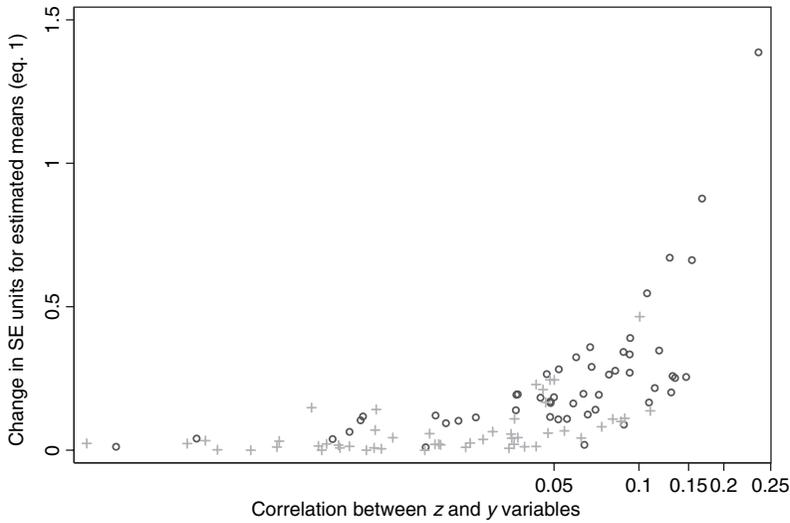
$$||\rho(\hat{p}_{xz}, y)| - |\rho(\hat{p}_x, y)|| \tag{2}$$

with  $\hat{p}_{xz}$  symbolizing the predicted values from either the response propensity models or the weighting class adjustments including both traditional covariates  $x$  and the new adjustment variable  $z$ . Likewise,  $\hat{p}_x$  represents the predicted values from response propensity models or weighting class adjustments values with only the traditional adjustment covariates  $x$ .

Using the difference in the correlation can also be motivated in a different way. Although not a measure of the non-response bias itself, these correlations nevertheless give a scale-free assessment of relative influence on the non-response bias. Under a stochastic model for survey non-response (Bethlehem, 2002), the non-response bias of an unadjusted respondent mean  $\bar{y}$  can be expressed as

$$\text{bias}(\bar{y}) = \frac{\text{cov}(p, y)}{\bar{p}}, \tag{3}$$

i.e. the covariance between the response probabilities and the survey outcome variable over the average of the response probabilities which equates to the response rate in a given survey.



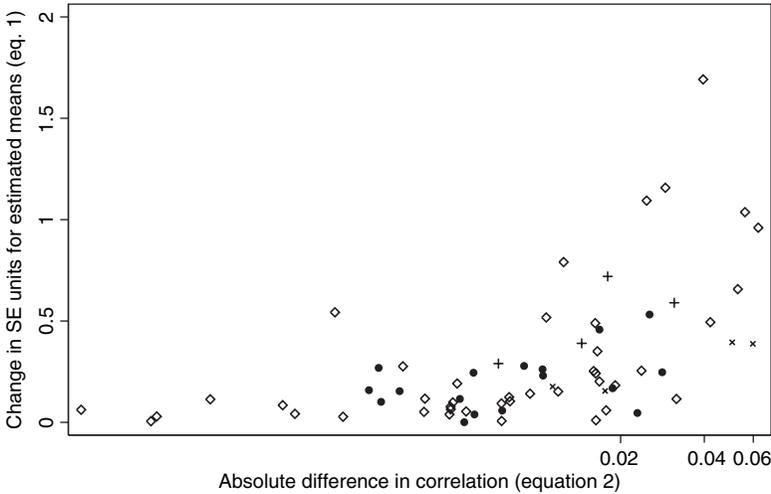
**Fig. 2.** Absolute change in standard error units for estimated means and estimated percentages of outcome variables  $y$  listed in Table 4 plotted against the correlation between interviewer observations and outcome variables for the ESS (log-scale): +,  $z$ -variable presence of litter; O,  $z$ -variable multiunit housing

Although non-response bias of the unadjusted respondent mean can be expressed as a covariance between  $p$  and  $y$ , the covariance is on the scale of the survey outcome variable. As such, comparing across  $y$ -variables within the same survey is difficult. To make it scale free, we divide both sides of equation (3) by the standard deviation of  $y$ . For ease of interpretation, we then re-express the scale-free estimate of non-response bias as the correlation between the response probability and  $y$  times the coefficient of variation of  $p$ :

$$\frac{\text{bias}(\bar{y})}{\text{SD}(y)} = \frac{\text{cov}(p, y)}{\text{SD}(y)\bar{p}} = \frac{\text{cov}(p, y)}{\text{SD}(y)\text{SD}(p)} \frac{\text{SD}(p)}{\bar{p}} = \rho(p, y) \text{CV}(p). \quad (4)$$

Within a survey the coefficient of variation of  $p$  is fixed. The correlation permits the relationship between response probability and  $y$  to be expressed in standard deviation units, thus making the results more comparable. Previous investigations of the effects of weighting on survey estimates (Kish, 1965) focus on the second term,  $\text{CV}(p)$ . We focus on the first term. All things being equal, stronger associations between  $y$  and  $p$  will lead to increased non-response bias of the unadjusted respondent mean within a survey.

Fig. 3 shows the absolute change in the estimate weighted with the old and new adjustment weights (see equation (1)) for the four surveys for which traditional non-response adjustment was available (the ANES, NSFG, UMTRI and MEPS). The  $y$ -axis shows the absolute change in the weighted estimated means in standard error units. The  $x$ -axis displays the absolute difference in the correlations between the estimated response probability and  $y$  including and excluding the new adjustment variable  $z$  (see equation (2)). Similarly to Fig. 2 we see an increase in the change of the weighted estimate with increasing difference in the correlation of the estimated response probabilities and the outcome variables. The range of values on the  $x$ -axis is considerably lower than that in Fig. 2. Remember that the newly added adjustment variable  $z$  is only one of many variables included in the models that are used to estimate the response probability. This is particularly true for the NSFG and MEPS. For many of the survey variables, adding the new adjustment variable  $z$  to the propensity model led to little change in the correlation of the estimated response probability with the survey outcomes as displayed on the  $x$ -axis. However,



**Fig. 3.** Absolute change in standard error units for estimated means and estimated percentages (see equation (1)) of the outcome variables measured in Tables 2, 3, 5 and 6, plotted against the absolute difference (log-scale) in correlation of the estimated response probability and the outcome variables for the UMTRI (+), MEPS (x), ANES (◇) and NSFG (●) studies

we do see for some variables shifts in the mean estimates of approximately 1 standard error of the estimate in magnitude. For most surveys adding  $z$  to the propensity model increases the strength of  $\rho(\hat{p}_{xz}, y)$  compared with  $\rho(\hat{p}_x, y)$  for some survey outcome variables. Yet this was not universally true.

**4.2. Direction of change and effect on variances**

Establishing the potential of  $z$  to change the estimated mean is only a first step. Survey researchers will ultimately be interested if using those  $z$ -variables will remove non-response bias and what the effect of the new weighting adjustment is on the estimated variances.

Adjustment weights can increase the estimated variance of survey estimates in addition to affecting the point estimate when the weights are only weakly correlated with the survey outcome (Little and Vartivarian, 2003, 2005). We are therefore interested in the trade-off between non-response bias reduction (due to a potentially better non-response adjustment) and the potential increase in variance.

For three of the studies (the ESS, NSFG and ANES) we cannot evaluate the effect of the shift in terms of non-response bias due to the absence of true scores or strong assumptions regarding a potential non-response bias. Thus, a non-response bias examination will be done for two of the five surveys: the MEPS and UMTRI study. For the MEPS, a non-response bias assessment is possible by using the NHIS frame data. From the NHIS, outcome variables are available for both MEPS respondents and non-respondents. For the UMTRI study, we assume that concerns about revealing socially undesirable information increase the probability of refusing to participate in the UMTRI survey. Hence, an increase in the non-response-weighted estimates of the undesirable behaviours will be taken as evidence for a shift in the appropriate direction.

The mean-square error (MSE) reflects the contribution of both variance and bias on the estimate. We can obtain an estimate for the MSE in the MEPS by defining the bias in terms of a target value rather than the expected value (Wun and Ezzati-Rice, 2007). For the MEPS the target value is the value that is estimated for the full ( $f$ ) MEPS sample (respondent and

non-respondents) using the MEPS base weight ( $\hat{y}_{w_f}$ ), i.e. before any non-response adjustment. Here we use the square root of the MSE divided by the target value ( $\hat{y}_{w_f}$ ), obtaining a relative root-mean-square error rRMSE to compare the different sets of non-response weights. Each bias term is estimated as the difference between the non-response-adjusted estimate  $\hat{y}_{w_x}$  and  $\hat{y}_{w_{xz}}$ , and the target value  $\hat{y}_{w_f}$ . Each variance term is defined as the non-response-adjusted estimate,  $\hat{y}_{w_x}$  and  $\hat{y}_{w_{xz}}$ , centred at the target value  $\hat{y}_{w_f}$ . Equation (5) presents rRMSE for  $\hat{y}_{w_{xz}}$ , and there is a corresponding expression for  $\hat{y}_{w_x}$ :

$$\text{rRMSE} = \frac{\sqrt{\{\text{bias}^2(\hat{y}_{w_{xz}}) + \widehat{\text{var}}(\hat{y}_{w_{xz}})\}}}{\hat{y}_{w_f}}, \tag{5}$$

For the UMTRI study we do not know the absolute magnitude of the non-response bias; thus we cannot estimate the change in MSE. However, we can report a ratio  $R$  of the variances: the variance of the estimated mean by using the traditional covariates for the non-response weighting adjustment ( $\widehat{\text{var}}(\hat{y}_{w_x})$ ) and the variance by using both traditional and new  $z$ -variables in the adjustment procedure ( $\widehat{\text{var}}(\hat{y}_{w_{xz}})$ ):

$$R = \widehat{\text{var}}(\hat{y}_{w_x}) / \widehat{\text{var}}(\hat{y}_{w_{xz}}). \tag{6}$$

If the new adjustment variables are highly correlated with the survey outcome, we would expect the variance in the denominator ( $\widehat{\text{var}}(\hat{y}_{w_{xz}})$ ) to be smaller than the variance in the numerator ( $\widehat{\text{var}}(\hat{y}_{w_x})$ ), and thus  $R$  would be greater than 1.

Because the weights are estimated, they also have sampling variability. To account for the sampling variability of the weights, we use resampling methods. We take 100 draws from a multivariate normal distribution with  $\text{MN}\{\hat{\beta}, \text{cov}(\hat{\beta})\}$ , where the vector of regression coefficients  $\hat{\beta}$  and  $\text{cov}(\hat{\beta})$  are taken from the logit models that were discussed earlier (see Table 2). For each draw, we estimate a new response probability  $\hat{p}_i$  and a non-response adjustment weight  $1/\hat{p}_i$ , and we re-estimate the weighted means and variances by using each of these 100 propensity weights as described above. We then estimate the mean of the variances and the variance of the means to calculate the unconditional variance  $v$ , which is  $v = E\{\widehat{\text{var}}(\hat{y}|\hat{\beta})\} + \widehat{\text{var}}\{E(\hat{y}|\hat{\beta})\}$ .

Table 8 presents the results for the MEPS. In Table 8 means and percentages displayed in the third column ( $\hat{y}_{w_x}$ ) are estimated by using the traditional non-response adjustment weight and the selection weight. The estimates in the fifth column ( $\hat{y}_{w_{xz}}$ ) add the two new health-related adjustment variables  $z$  to the non-response adjustment weight. The RMSE is displayed next to each non-response-weighted estimate. As a reminder, for the MEPS, survey values (from the source sampling frame) exist for both respondents and non-respondents. Hence, the non-response-weighted estimates for the respondents can be compared with the target value ( $\hat{y}_{w_f}$ ). Non-response-weighted estimates are better if they are closer to the target value.

Adding the health-related  $z$ -variables to the non-response adjustment moves the resulting estimate closer to the target value for all four outcome variables, i.e., for all four outcome variables, we observe that *adding* health-related covariates in the dwelling unit level non-response adjustment results in estimates with smaller non-response bias and reduced variation. As a consequence, the MEPS now includes health-related variables in the dwelling unit level non-response adjustment.

As argued above, for the UMTRI study concerns about revealing socially undesirable information may increase non-response and thus lead to non-response bias for survey outcome variables that are related to the undesirable information. Thus if the new  $z$ s are useful adjustment variables we would expect increases in the non-response-weighted estimates, compared

**Table 8.** Unweighted and weighted estimates for the MEPS

Estimate	Target value $\hat{y}_{wf}$	Old adjustment		New adjustment	
		$\hat{y}_{wx}$	rRMSE (%)	$\hat{y}_{wxz}$	rRMSE (%)
% with no limitations in daily activities	74.64	72.74	2.76	73.05	2.38
Mean dollar index	6618	6712	1.82	6682	1.48
% with no doctor visit in past 12 months	67.23	65.67	2.53	65.79	2.37
% with no barrier to healthcare owing to cost	85.28	84.28	1.36	84.37	1.27

**Table 9.** Unweighted and weighted estimates for the UMTRI survey

Estimate	Unadjusted $\hat{y}_w$	Adjusted		Standard errors $\times 10^{-3}$		R
		$\hat{y}_{wx}$	$\hat{y}_{wxz}$	SE( $\hat{y}_{wx}$ )	SE( $\hat{y}_{wxz}$ )	
Proportion with licence revoked	18.80	19.43	19.85	5.74	5.88	0.98
Mean number of traffic tickets	0.35	0.36	0.37	0.119	0.121	0.98
Proportion never wearing seat belt on local travels	5.13	5.25	5.38	3.22	3.32	0.97
Proportion never wearing seat belt on long travels	4.28	4.39	4.47	2.96	3.04	0.97

with those estimated only with the selection weight ( $\hat{y}_w$ ). In Table 9, the column headed with  $\hat{y}_{wx}$  displays UMTRI estimates by using the traditional non-response adjustment weight and the selection weight, and  $\hat{y}_{wxz}$  indicates estimates for which the new adjustment variables  $z$  are included when forming the non-response adjustment weight.

For example, as we see in Table 9, the unadjusted estimate for the number of young adults having their licence was revoked is 18.80%. When applying the original non-response adjustment weight formed by cells age and gender the estimated proportion of young adults whose licence was revoked is 19.43%. When the five new  $z$ -variables are added to the existing covariates in the adjustment model the estimated proportion of drivers with their licence revoked increases yet again to 19.85%. Thus the non-response-weighted estimates change in the expected direction. The estimated total variances are small. The variance for the estimate by using the full set of  $z'$  and  $x$  adjustment variables is only slightly larger than the existing adjustment by using only the  $x$ -variables; as a result, the ratio  $R$  (see equation (6)) is slightly less than 1 in the last column. In sum, for the two surveys in which a non-response bias assessment was possible, the weights that are formed with the new adjustment variables reduced non-response bias at no great expense of increased variance.

### 5. Discussion

This paper explores the use of auxiliary variables that are in part proxy measures of survey variables in post-survey adjustments. We used the NSFG, the MEPS, the ANES, the ESS and the UMTRI survey as examples. For most of the surveys that are evaluated here, the auxiliary

variables are moderately correlated with survey outcomes, and less so with the response indicator. Overall the correlations are weak and do not show the strength that is needed for successful non-response adjustment as described by Little and Vartivarian (2005).

We compared unweighted and weighted estimates under various non-response models. Including the proxy measures in the weight construction led to a larger shift in the survey estimates than when they are not included. In addition, the extent of change in the survey estimates by using the new weights is related to the strength of the associations between the survey outcomes and the estimated response probabilities when using the new auxiliary variables—larger shifts in estimates are observed with moderate correlations and smaller effects with lower correlations. Owing to the weak correlations between most proxy measures and the response indicator, we did not observe any apparent changes in the variance of the estimates.

We saw strong variability across survey items (even within a single survey) for the bivariate correlations as well as in the effects of the adjustment on the estimated means. For any given correlational strength between the proxy measure and the response indicator, we also saw large variation in the correlation of that same proxy variable and the survey outcomes. Consequently, relying on one or a few variables in the adjustment may not necessarily improve weighted estimates across all outcomes. Although a strong predictor of response (to the survey request) can be important for the weighting adjustment, care should be taken in evaluating the weighting models. Predictors of the survey outcome variables that are also predictors of response should be deliberately sought out and included in adjustment models, in addition to the traditional predictors of survey participation.

This paper did not consider the measurement error properties of the interviewer observations and record variables. We made a simplistic assumption that there is no measurement error in those variables. Of course, this assumption is debatable in the real world. Future research is needed to examine the effect of the potential measurement error in auxiliary variables on survey estimates and on the bias–variance trade-off. Although it will be difficult to do so, research is also needed on the presence and effect of selective measurement error, e.g. if measurement error in the auxiliary variables is correlated with response.

This paper focused only on changes in estimates of central tendencies (i.e. weighted mean estimates or weighted estimated proportions of binary variables). It is conceivable that larger shifts can be observed in other estimates such as quantiles. Also, an ideal proxy variable for a successful adjustment variable for a point estimate does not necessarily translate to the ideal variable when interest is in estimated regression coefficients or subgroup analyses.

In this paper an assessment of reduction of non-response bias was only possible for the MEPS, and with some assumptions for the UMTRI study. Very few surveys have a record base for their frame that contains information on both respondents and non-respondents, and those are often limited to behavioural items. Likewise very few surveys are part of a panel study where values for non-respondents are available from prior waves. However, the strength of this paper is that we have examined alternative variables over multiple surveys with multiple recruitment protocols from multiple organizations on multiple topics.

The challenge of weighting adjustment, for survey researchers and practitioners, lies in the search for an appropriate set of auxiliary variables that are predictive of both response probabilities and survey variables of interest. We encourage survey researchers to engage actively in identifying an appropriate set of auxiliary variables in developing non-response adjustment weights. This should include identifying measures at the design stage that can be obtained on both respondents and non-respondents and that are good proxy variables for one or multiple survey variables. In the past, attention was often focused on finding variables that are associated with response although small  $R^2$ -statistics are very common in response propensity models

(Janson, 2003; Nicoletti and Peracchi, 2005; Feskens *et al.*, 2007). The results of this paper show that a renewed focus on correlates of the key survey outcome variables is warranted. An avenue that is worth exploring is statistics derived from call record data or other types of paradata that were not discussed here.

## Acknowledgements

The paper resulted from a research seminar run by Robert Groves and Trivellore Raghunathan at the Joint Program in Survey Methodology and the Michigan Program in Survey Methodology. We thank Katharine Abraham, Jill Dever, Stephanie Eckman, Jenna Fulton, Susan McCulloch, Elizabeth Stuart, Richard Valliant, the Joint Editor and three reviewers for comments and helpful suggestions, and we thank Lap-Ming Wun of the Agency for Healthcare Research and Quality for computer programming assistance. The views that are expressed in this paper are those of the authors and carry no official endorsement by the Department of Health and Human Services or the Agency for Healthcare Research and Quality.

## References

- Abraham, K. G., Maitland, A. and Bianchi, S. M. (2006) Nonresponse in the American Time Use Survey: who is missing from the data and how much does it matter? *Publ. Opin. Q.*, **70**, 676–703.
- Atrostic, B., Bates, N. and Silberstein, A. (2001) Nonresponse in U.S. government household surveys: consistent measures, recent trends, and new insights. *J. Off. Statist.*, **17**, 209–226.
- Bates, N., Dahlhamer, J. and Singer, E. (2008) Privacy concerns, too busy, or just not interested: using doorstep concerns to predict survey nonresponse. *J. Off. Statist.*, **24**, 591–612.
- Bethlehem, J. (2002) Weighting nonresponse adjustments based on auxiliary information. In *Survey Nonresponse* (eds R. Groves, D. Dillman, J. Eltinge and R. Little), pp. 275–287. New York: Wiley.
- Bingham, C. (2007) *Crash Risk among Teen Drivers: Identification and Prediction of Excess Risk*. Ann Arbor: University of Michigan Transportation Research Institute.
- Center for Political Studies (2004) The 2004 National Election Study data set. Center for Political Studies, University of Michigan, Ann Arbor.
- Cohen, J. W. (1997) *Design and Methods of the Medical Expenditure Panel Survey Household Component*. Rockville: Agency for Health Care Policy and Research.
- Copas, A. J. and Farewell, V. T. (1998) Dealing with non-ignorable non-response by using an ‘enthusiasm-to-respond’ variable. *J. R. Statist. Soc. A*, **161**, 385–396.
- Couper, M. P. (1997) Survey introductions and data quality. *Publ. Opin. Q.*, **61**, 317–338.
- Couper, M. P. (1998) Measuring survey quality in a CASIC environment. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 41–49.
- Couper, M. P., Singer, E. and Kulka, R. A. (1998) Participation in the 1990 decennial census—politics, privacy, pressures. *Am. Polit. Q.*, **26**, 59–80.
- Curtain, R., Presser, S. and Singer, E. (2005) Changes in telephone survey nonresponse over the past quarter century. *Publ. Opin. Q.*, **69**, 87–98.
- De Leeuw, E. and DeHeer, W. (2002) Trends in household survey nonresponse: a longitudinal and international comparison. In *Survey Nonresponse* (eds R. Groves, D. Dillman, J. Eltinge and R. Little), pp. 41–54. New York: Wiley.
- Dillman, D., Eltinge, J., Groves, R. and Little, R. (2002) Survey nonresponse in design, data collection and analysis. In *Survey Nonresponse* (eds R. Groves, D. Dillman, J. Eltinge and R. Little), pp. 3–26. New York: Wiley.
- Ezzati-Rice, T. M., Rohde, F. and Greenblatt, J. (2008) *Sample Design of the Medical Expenditure Panel Survey, 1998–2007*. Rockville: Agency for Healthcare Research and Quality.
- Feskens, R., Hox, J., Lensvelt-Mulders, G. and Schmeets, H. (2007) Nonresponse among ethnic minorities: a multivariate analysis. *J. Off. Statist.*, **23**, 387–408.
- Groves, R. M. (2006) Nonresponse rates and non-response bias in household surveys. *Publ. Opin. Q.*, **70**, 646–675.
- Groves, R. and Couper, M. (1998) *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R. and McGonagle, K. (2001) A theory-guided interviewer training protocol regarding survey participation. *J. Off. Statist.*, **17**, 249–266.
- Groves, R. and Peytcheva, E. (2008) The impact of non-response rates on non-response bias: a meta-analysis. *Publ. Opin. Q.*, **72**, 167–189.

- Groves, B., Wagner, J. and Peytcheva, E. (2007) Use of interviewer judgments about attributes of selected respondents in post-survey adjustment for unit nonresponse: an illustration with the National Survey of Family Growth. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 3428–3431.
- Janson, H. (2003) Influences on participation rate in a national Norwegian child development screening questionnaire study. *Acta Paed.*, **92**, 91–96.
- Johnson, T., Cho, Y., Campbell, R. and Holbrook, A. (2006) Using community-level correlates to evaluate nonresponse effects in a telephone survey. *Publ. Opin. Q.*, **70**, 704–719.
- Kalton, G. and Flores-Cervantes, I. (2003) Weighting methods. *J. Off. Statist.*, **19**, 81–97.
- Kish, L. (1965) *Survey Sampling*. New York: Wiley.
- Kreuter, F., Lemay, M. and Casas-Cordero, C. (2007) Using proxy measures of survey outcomes in post-survey adjustments: examples from the European Social Survey (ESS). *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 3142–3149.
- Lepkowski, J., Mosher, W., Davis, K., Groves, R., van Hoewyk, J. and Willem, J. (2006) National Survey of Family Growth, Cycle 6: sample design, weighting, imputation, and variance estimation. *Vitl Hlth Statist. Ser.* 2, no. 142, 1–82.
- Little, R. (1986) Survey nonresponse adjustments for estimates of means. *Int. Statist. Rev.*, **54**, 139–157.
- Little, R. and Rubin, D. (2002) *Statistical Analysis with Missing Data*. Hoboken: Wiley.
- Little, R. J. and Vartivarian, S. (2003) On weighting the rates in non-response weights. *Statist. Med.*, **22**, 1589–1599.
- Little, R. and Vartivarian, S. (2005) Does weighting for nonresponse increase the variance of survey means? *Surv. Methodol.*, **31**, 161–168.
- Lynn, P. (2003) PEDAKSI: methodology for collecting data about survey non-respondents. *Qual. Quant.*, **37**, 239–261.
- Nicoletti, C. and Peracchi, F. (2005) Survey response and survey characteristics: microlevel evidence from the European Community Household Panel. *J. R. Statist. Soc. A*, **168**, 763–781.
- Peytchev, A. and Olson, K. (2007) Using interviewer observations to improve nonresponse adjustments: NES 2004. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 3364–3371.
- Rogers, A., Murtaugh, M. A., Edwards, S. and Slattery, M. L. (2004) Contacting controls: are we working harder for similar response rates and does it make a difference? *Am. J. Epidem.*, **160**, 85–90.
- Rosenbaum, P. and Rubin, D. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55.
- Schnell, R. (1997) *Nonresponse in Bevölkerungsumfragen*. Opladen: Leske and Budrich.
- Wun, L.-M. and Ezzati-Rice, T. (2007) Assessment of the impact of health variables on nonresponse adjustment in the medical expenditure panel survey (MEPS). *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 2857–2864.
- Yan, T. and Raghunathan, T. (2007) Using proxy measures of the survey variables in post-survey adjustment in a transportation survey. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 3349–3355.