An Empirical Study of Nonresponse Adjustment Methods for the Survey of Doctorate Recipients[1]

Fan Zhang[1] and Stephen Cohen[1]
Donsig Jang[2], Amang Sukasih[2], and Sonya Vartivarian[2]
[1]National Science Foundation, Division of Science Resources Statistics,
4201 Wilson Blvd., Suite 965, Arlington, VA 22230
[2]Mathematica Policy Research. Inc., 600 Maryland Ave., SW, Suite 550,
Washington DC 20024-2512

Abstract:
In the Survey of Doctorate Recipients at National Science Foundation, adjustments were made to the design weights to compensate three types of missing units: un-located, eligibility-unknown, and refusals. In search of sensible adjustment methods, we conducted a simulation study to evaluate five different adjustment factors: the inverse of unweighted response rate in weighting cell formed by cross-tabulation of significant main effects; the inverse of weighted response rate in weighting cell formed by deciles of predicted propensity values estimated by logistic regression; the inverse of unweighted response rate in weighting cell formed by deciles of predicted propensity values estimated by logistic regression; the inverse of individual predicted propensity values from an unweighted logistic regression model; and the inverse of individual predicted propensity values from a weighted logistic regression model.

Key Words:
Survey of Doctorate Recipients, nonresponse adjustment, covariate weighting class, propensity weighting class, inverse propensity score

1. Introduction

The Science and Engineers Statistical Data System (SESTAT) at National Science Foundation (NSF) is an integrated data system collecting information about employment, education, and demographic characteristics of scientists and engineers in the United States. The data are collected from three national surveys that are independently conducted and cover different segments of the SESTAT population: the National Survey of College Graduates (NSCG), the National Survey of Recent College Graduates (NSRCG), and the Survey of Doctorate Recipients (SDR). NSCG represents all individuals in the U.S. at the time of the decennial census with a bachelor’s degree or higher, NSRCG represents persons with a bachelor’s or master’s degree in science and engineering from a U.S. institution earned since the decennial census, and SDR represents persons in the general U.S. population who have earned a doctorate in science or engineering from a U.S. institution. As a whole, SESTAT provides information on the entire U.S. population of scientists and engineers with at least a bachelor’s degree.

---

[1] This paper is intended to report exploratory results of research and analysis undertaken by the Division of Science Resources Statistics. Any opinions, findings, conclusions or recommendations expressed in this paper do not necessarily reflect the views of the National Science Foundation.

Like many other federal government surveys, NSF's SESTAT surveys have also suffered declining unit response rates in recent years. During the ten year period of time from 1993 to 2003, the overall unweighted response rates for NSCG, NSRCG and SDR declined from 80% to 63%, 85% to 67%, and 87% to 79% respectively.

Unit non-response, however, can happen for different reasons. Figure 1 show a typical SESTAT sample decomposition. A sampled person may become a non-respondent because we are unable to locate him/her. Or among the located cases, a person becomes non-respondent because we are unable to establish a contact with him/her and therefore this person has unknown eligibility to the survey. Or among the known eligible cases (especially those panel cases), a person becomes non-respondent simply because this person refuses to participate the survey.

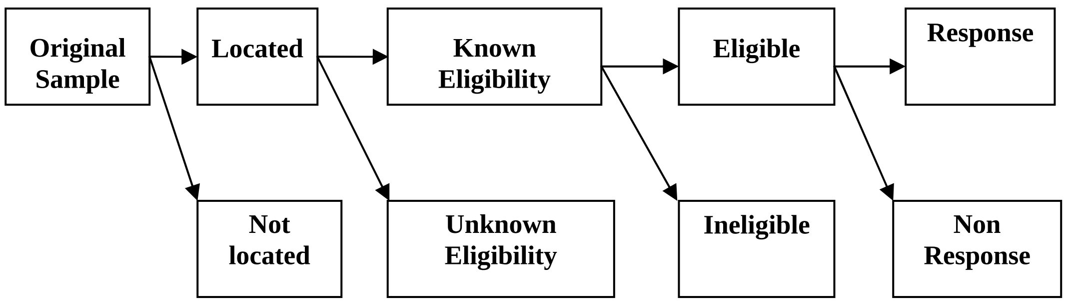


Figure 1: Sample Decomposition

High non-response rates are a serious problem because of the bias arising from the missing cases and from the potential differences between respondents and non-respondents. Various weight adjustment methods for handling non-response in sample surveys have been developed. Kalton and Kasprzyk 1986, Brick and Kalton 1996 provide a good review of these methods.

To better understand the current method used in SESTAT and compare it with alternative methods, we identified three types of weighting adjustment methods: covariate weighting class, propensity weighting class and inverse propensity score and conduct an empirical study to compare these three types of weighting adjustments.

The covariate weighting class method uses covariates to create weighting classes and then calculates adjustment factor for each weighting class as the non-response adjustment factor. In SESTAT, for example, base weights are adjusted by multiple covariate weighting class adjustment factors, e.g: non-locating, unknown eligibility, and non-response to compensate the unit non-response:

$$w_k = f_{1k} f_{2k} f_{3k} b_k ,$$

here, w_k is the nonresponse adjusted weight, b_k is base weight, f_{1k} is the inverse of weighted locating rate, f_{2k} is the inverse of weighted known eligibility rate, f_{3k} is the inverse of weighted response rate. All rates are calculated by weighting classes formed by frame variables. In our study, we first fit logistic regression models to identify the covariates that are associated with the response status and only use the significant covariates to create weighting cells.

The propensity weighting class method uses the estimated response propensity score to form the weighting classes and then calculates adjustment factor for each weighting class as the adjustment factor. In our study we first fit logistic regression models using frame variables to estimate the response probabilities of each case. Then group the estimated response probabilities into ten weighting classes using the decile of the estimated response probabilities, from the smallest to the largest and use the weighted weighting class response rates as the adjustment factors.

The inverse propensity score method simply uses the estimated individual propensity scores as the adjustment factor for each case. In our study, we use the fitted individual propensity score estimated from logistic regression models.

Our empirical study was conducted to two SESTAT component surveys: NSRCG and SDR. This paper only deals with the SDR portion.

Section 2 gives a description of SDR and non-response adjustment used in 2006 SDR. Section 3 is about the preliminary study that identifies the modeling options and weighting adjustments for the simulation study. Section 4 describes the design of the simulation study. Section 5 is the results of the simulation. Section 6 is a summary.

2. 2006 Survey of Doctor Recipients

The 2006 Survey of Doctorate Recipients (SDR) is a longitudinal survey of U.S. residents with U.S.-granted doctoral degrees in science and engineering (S&E). The target population of the 2006 SDR is individuals who received a doctoral degree in an S&E field from a U.S. institution, was 75 year or younger on April 1, 2006, and lives in the U.S. in a non-institutionalized setting on April 1, 2006.

The sampling frame of SDR is a list of all U.S.-granted research doctorates in a database called Doctorate Records File (DRF). DRF is a cumulative database of research doctorate recipients from U.S. institutions collected by an annual census called Survey of Earned Doctorates (SED) conducted by National Science Foundation.

The 2006 SDR used a stratified sample design. In each stratum, the sample comprises two sub samples selected independently from two sub frames: the 2006 SDR old cohort frame and the 2006 SDR new cohort frame. The 2006 SDR new cohort frame is developed from a list of all research doctoral degrees rewarded between July 1, 2003 and June 30, 2006. The 2006 SDR old cohort frame is developed from the 2003 SDR original sample (including both respondents and non-respondents). The 2003 SDR original sample was selected from research doctoral degrees rewarded before July 1, 2003. Therefore the 2003 SDR original sample serves as the first phase sample of 2006 SDR old cohort sampling. The 2006 SDR new cohort frame is developed from 2004 and 2005 research doctorate recipients listed in 2004 and 2005 SED.

Before sample selection, both old cohort and new cohort are stratified into 164 strata defined by the cross of the three variables: demographic group, gender, and degree field. An equivalent of at least 60 cases (new cohort and old cohort combined) is allocated to each stratum as the minimum stratum sample size. Also ten critical analysis domains are identified and a domain supplemental allocation is set to support analysis on these domains. Total 38,027 cases are allocated to the old cohort sample and 4,928 cases are allocated to the new cohort sample. An iterative sample selection procedure is used to select the sample. For detailed 2006 SDR sample design and implementation information, see Yang et al 2007.

The nonresponse adjustment for the 2006 SDR is performed in two steps: first adjust for the cases with unknown eligibility and then adjust for non-responding cases. Here, the cases with unknown eligibility are those were never located or were never contacted during the data collection period. In other words, in 2006 SDR, the adjustment for unable to locate and the adjustment for unknown eligibility were combined into one adjustment. The SDR sampling strata are used as the nonresponse weight adjustment cells. The following is a summary of the adjustments methods from Yang 2007:

k : The adjustment cell;
ER_k : The sum of the base weight (w_{1i}) over all eligible respondents of cell k ;
EN_k : The sum of the base weight (w_{1i}) over all eligible non-respondents of cell k ;
UN_k : The sum of the base weight (w_{1i}) over all unknown eligibility cases of cell k ;
IN_k : The sum of the base weight (w_{1i}) over all known ineligible cases of cell k ;

To adjust the unknown eligibility, the unknown eligibility adjustment factor for cell k is computed as:

$$f_{1k} = \frac{ER_k + EN_k + IN_k + UN_k}{ER_k + EN_k + IN_k}$$

Through this adjustment, the base weight carried by the unknown eligibility cases is distributed proportionately to cases with known eligibility. The eligibility adjusted weight is

$$w_{2i} = w_{1i} * f_{1k}$$

The weight for the eligible respondents is further adjusted to compensate for eligible non-respondents. The nonresponse adjustment factor for eligible completes in cell k is calculated as,

$$f_{2k} = \frac{ER_k^{'} + EN_k^{'}}{ER_k^{'}}$$

where $ER_k^{'}$ is the sum of the eligibility adjusted weight (w_{2i}) over all eligible respondents of cell k ; and $EN_k^{'}$ is the sum of the eligibility adjusted weight (w_{2i}) over all eligible non-respondents of cell k . See Yang (2007) for detailed description on 2006 SDR weighting method. The nonresponse adjusted weight is

$$w_i = f_{2k} w_{2i}$$

The final adjustment cells are slightly different from the sampling strata due to cell collapsing to avoid small cells and large adjustment factors. For the unknown eligibility adjustment, a cell is collapsed with another cell if there are less than 20 cases with known eligibility. For the non-response adjustment, a cell is collapsed with another cell if there are less than 20 eligible respondents. For both adjustments, a cell would be collapsed with another cell if the adjustment factor is grater than 2. It turns out all collapsing carried out is due to small cell size and no cell is collapsed because of a large adjustment factor.

3. Preliminary Analysis

As we have seen in section 2, the sampling frame of SDR is the Doctorate Records File (DRF). DRF is a cumulative database of all research doctorates rewarded in the U.S. institutions, collected by an annual census called Survey of Earned Doctorates (SED). Therefore, in addition to the sampling variables used for SDR sample design, there are many other auxiliary variables in DRF which are useful for this study. The variables used in this investigation are listed in Table 3.

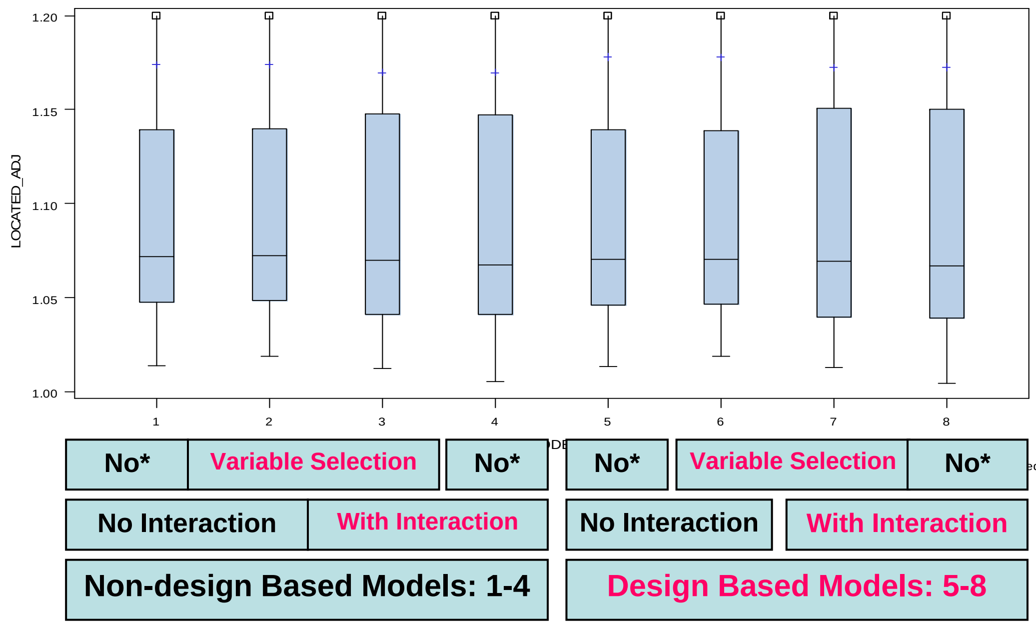
All three types of adjustment methods mentioned above require logistic regression modeling. Therefore before we conduct the simulation study, we conducted empirical investigation to assess the various modeling options to be used in the simulation. In particular, we compared the design-based modeling (taking the weighting and sample design feature into account) vs. the non-design based modeling, model with main effects only vs. a model with main effects and interaction terms, and a full model versus a reduced model based on stepwise variable selection. Therefore, eight different combination of modeling options were compared. We calculated the response propensity scores under various modeling options and compared the adjustment factors and the weights at each adjustment stage of nonresponse. Table 1 lists all eight different options we considered in this study.

Table 1: Eight Different Options of Fitting Logistic Regression Model

Model	**Design Based**	**Interaction Terms**	**Variable Selection**
1	N	N	N
2	N	N	Y
3	N	Y	Y
4	N	Y	N
5	Y	N	N
6	Y	N	Y
7	Y	Y	Y
8	Y	Y	N

Figure 2 is the clipped box plot of estimated response propensity scores for the locating indicator. Models 1-4 are non-design based models and model 5-8 are design based. Models 1, 2 and 5, 6 do not include any interaction terms while models 3, 4 and 7, 8 include the significant interaction terms identified by CHAID. Model 1, 4, 5, 8 are full models without stepwise variable selection and models 2, 3, 6, 7 are models with stepwise variable selection.

Figure 2 shows that design based or non-design based model fitting doesn't make much difference on the distribution of the propensity scores for SDR. This is because all sampling variables, i.e. the stratification variables, are considered in all the models. Figure 2 also shows with and without interaction terms indeed make differences. The effect of interaction terms can be further seen from Table 2. The Hosmer and Lemeshow Goodness-of-Fit statistics in Table 2 show interaction terms improve model fitting. The full models (1, 4, 5, 8) have better model fitting but also have larger variation among the propensity scores.



*: No= Without Variable Selection

Figure 2 - Box Plot of Estimated Response Propensity Scores for Locating Indicator

Table 2 - Hosmer and Lemeshow Goodness-of-Fit Statistics

	NON-LOCATING			ELIGIBILITY UNKNOWN		
Model	Chi-Square	DF	Pr > ChiSq	Chi-Square	DF	Pr > ChiSq
1	66.34	8	<.0001	12.25	8	0.1403
2	71.16	8	<.0001	10.23	8	0.249
3	13.3	8	0.1019	9.02	8	0.3404
4	8.98	8	0.3438	5.37	8	0.717

In summary, preliminary empirical analysis suggests either design based or non-design based modeling can be used for simulation study for our case. But we decide to carry both options to simulation for further investigation. Although interaction models are preferable to without interaction models, later in our simulation process we realized that it seems there is no guarantee that the 0% and 100% response rate cells can always be resolved for logistic regression modeling. This problem may happen to any collapsing procedure if one main effect appears in multiple interactions and the collapsing is performed in a sequential manner. Another issue with interaction terms during simulation is colinearity. In our case, interaction terms almost always cause colinearity in simulation. Therefore, we decided not to include the interaction terms in the simulation. Accordingly, we dropped the interaction terms in the models that were used to generate missing indicators in the simulation. Finally, we also decided to use stepwise variable selection in the simulation.

4. Simulation Design

Based on the preliminary analysis results, we identified the following adjustment methods for the simulation study.

Simulation Method 1 (SM1). The inverse of unweighted response rate of covariate weighting class formed by significant main effects

Simulation Method 2 (SM2). The inverse of unweighted response rate of decile weighting class formed by the deciles of predicted propensity values from non-design based modeling

Simulation Method 3 (SM3). The inverse of weighted response rate of decile weighting class formed by the deciles of predicted propensity values from non-design based modeling

Simulation Method 4 (SM4). The inverse of individual predicted propensity values from non-design based modeling

Simulation Method 5 (SM5). The inverse of individual predicted propensity values from design-based modeling

We used 2006 SDR final sample (respondents only) as a full sample. When weighted by the final weights, this set of respondents represents the population of SDR. This data serve as the benchmark when evaluating simulated data. Given this full sample, we generated unit missing indicators and calculate adjustment factors for each of the five adjustment methods mentioned above and adjust the remaining non-missing accordingly. This process is performed at three levels: Locating, Eligibility and Response. This process is repeated 1000 times.

2006 SDR response rate over three stages (Locating, Eligibility, and Response) was about 78%. In simulation, we use over-all response rate of 73%, 78% and 83%. The third stage (Response) response rate was also very high (97%). Therefore, this stage adjustment is not performed in the simulation. To reach combined rates of 73%, 78% and 83% in two stages, we set the Eligibility stage rates at 91% and adjust the Locating stage rates accordingly.

Three different nonresponse mechanisms are used to generate missing indicators in this simulation:

Missing Completely at Random (MCAR): In MCAR, the missingness does not depend on any variables, neither covariate nor survey outcome variable. The missing probability therefore is a constant that equals to the preset overall response rates at different stages mentioned above.

Missing at Random (MAR): In MAR, the missingness depends on observed values of covariates. That is, the probability of unit missing differs from case to case depending on the values of their covariates. Three variants of MAR were used as the response propensities in our simulation:

> ***MAR1.*** Missing probability equals to the unweighted response rate in covariate weighting class constructed by cross-classification of significant variables identified using the original 2006 SDR sample that includes all cases - respondents and non-respondents;
>
> ***MAR2.*** Missing probability equals to the unweighted response rate in 10 decile weighting classes constructed by the estimated propensity scores using the original 2006 SDR sample that includes all cases - respondents and non-respondents;
>
> ***MAR3.*** Missing probability equals to individual estimated propensity score calculated through a design-based logistic regression using the original 2006 SDR sample that includes all cases - respondents and non-respondents.

Not Missing at Random (NMAR): In the NMAR, the missingness depends on the unobserved values of survey outcomes. In our simulation, we used a NMAR mechanism that depends on both survey outcome (salary) and covariate (working status). We assigned a lower missing probability to the working persons that have salary less than 100K.

We used SAS 9.2 to program SDR simulation. During the simulation, we realized the interaction terms used in the models inevitably cause collinearity among the variables and produce weighting cells with either 100% missing or 0% missing cases. It is impossible to handle this individually during the simulation. Therefore, we dropped the interaction terms in our baseline models when we produce the response probabilities in all missing mechanisms for all cases.

The simulation is intended to evaluate the result of weighting adjustment to account for possible nonresponse bias. Suppose $\hat{\theta}_0$ denotes the estimate calculated based on the 2006 SDR full sample, and $\hat{\theta}_r$ (r = 1, …, 1000) denotes the estimate calculated based on each simulated data. The following mean of differences is used to measure the magnitude of bias.

$$Bias\left(\hat{\theta}\right) = \frac{1}{R}\sum_{r=1}^{1000}\left(\hat{\theta}_r - \hat{\theta}_0\right)$$

For this evaluation, the survey estimates/statistics $\hat{\theta}$ to be compared are: total estimates, median salary, mean salary and population proportion estimates.

5. Simulation Results

In this section, we present a portion of the simulation results. For more output of the simulation study, please refer to Sukasih et al (2009). First we noticed (Table 4) that all adjustment methods (SM1-SM5) are unbiased under MCAR, as we expected. This is true for all types of estimates (total, median, mean and proportion). All adjustment methods (SM1-SM5) are biased under NMAR for all types of estimates. Our discussion bellow is all about MAR (MAR1-MAR3).

Overall speaking, under MAR all five adjustment methods (SM1-SM5) perform similarly. They are all unbiased or approximately unbiased. This can be also seen from Table 4. But there are a few exceptions. Figure 3 shows totals estimated under MAR1 (unweighted response rate in weighting cells formed by significant main effects). The middle horizontal dot line represents the average of 1000 total estimates, while the other two horizontal dot lines represent the confidence interval bounds for the mean of total estimates. Every little circle in the plot is a total estimate, from replicate 1 to 1000. Adjustment methods 1, 3, 4 and 5 (SM1, SM3, SM4, and SM5) are unbiased. Adjustment method 2 (SM2) (unweighted response rate in non-design based propensity cells) however, shows significant bias. Note that the missing data mechanism for MAR1 is based on interactions, while for the SDR, the models used in the response propensity estimation do not include interactions because of estimation issues. Since the propensity cell weighting adjustments (Methods 3 and 4) smooth the weights, it is suspected that these methods suffer from bias under MAR1 since these weighting methods use less information than the covariate cell and individual propensity weighting methods. Since the frame total is calculated as the sum of final weights across all respondents in the original data. It is not hard to show the weighting adjustment method 3 always produces adjusted weights that sum back to the true grand total. That explains why SM3 estimates have no error. For similar reason, SM5 produces highly efficient estimates.

This efficiency of SM3 and SM5, however, is only associated with the overall estimates. This can be seen from Figure 4. Figure 4 presents the plots for domain total estimates. We can see all methods are unbiased or approximately unbiased with similar variance.

Unequal probabilities of sample selection result in variability within the basic sampling weights. This variability results in larger variance of the survey estimate as compared to the variance from a self-weighted sample (with equal weight for everyone). For this reason, nonresponse weighting adjustment may add more variability within the respondent weights. We assess possible variance inflation due to the variability of the adjusted weights by the design effect of the weights, or $1+cv^2$. Here cv is the coefficient of variance of the weights. Table 3 lists this design effect of the weights for all adjustment methods at three different response rates. Under MCAR, neither adjustment methods nor the response rates have effect on the design effect. Under MAR or NMAR, however, the decile group methods (SM2 and SM3) have smallest design effects and individual propensity score methods (SM4 and SM5) have the largest design effects. But the differences are small especially when we notice the full sample weights have design effect of 1.114.

6. Summary

Based on our empirical and simulation investigation using the 2006 SDR, our main conclusions under MAR are as follows:

- Overall, all five methods considered for simulation are comparable.
- Weighting the response rate did not produce different results compared to the unweighted analysis, given that the weighting cells were constructed based on cross-tabulation of sampling variables and variables that strongly correlate with

- Weighting cells based on covariates can lead to issues with respect to small cell sizes and requires collapsing. Though in practice such collapsing strategy can be ad hoc, in our study simulation small cells were effectively handled with a systematic, automated method of collapsing.

- As the number of covariates including paradata available for use in the covariate cell creation increases, the covariate cell adjustment becomes a less desirable method because of sparseness and the increased need to collapse cells, ultimately limiting the ability to incorporate additional covariates.

- Alternatively, inverse propensity estimate adjustments maximize the utilization of auxiliary information and nonresponse bias reduction. However, there is often the concern that this may cause the most variable weights, and thus in turn, larger variances of estimates. In this study, more variable weights were observed, but the impacts on survey estimates were minimal in our SESTAT application due to a few cases having the largest weights. In practice, some of this weight variation may be dealt with via weight trimming.

- As an alternative to the covariate-based weighting class adjustments and the individual inverse propensity estimate adjustments, the hybrid technique—propensity cell adjustments are attractive in a sense that this method makes the response propensity distribution smooth (and thus making the weight variation less) while utilizing all

- Though we investigated three different response rates, our study indicates that weighting technique produces similar results under these three survey response rates. Given a weighting method, we did not observe major differences of the shape of weight distributions.

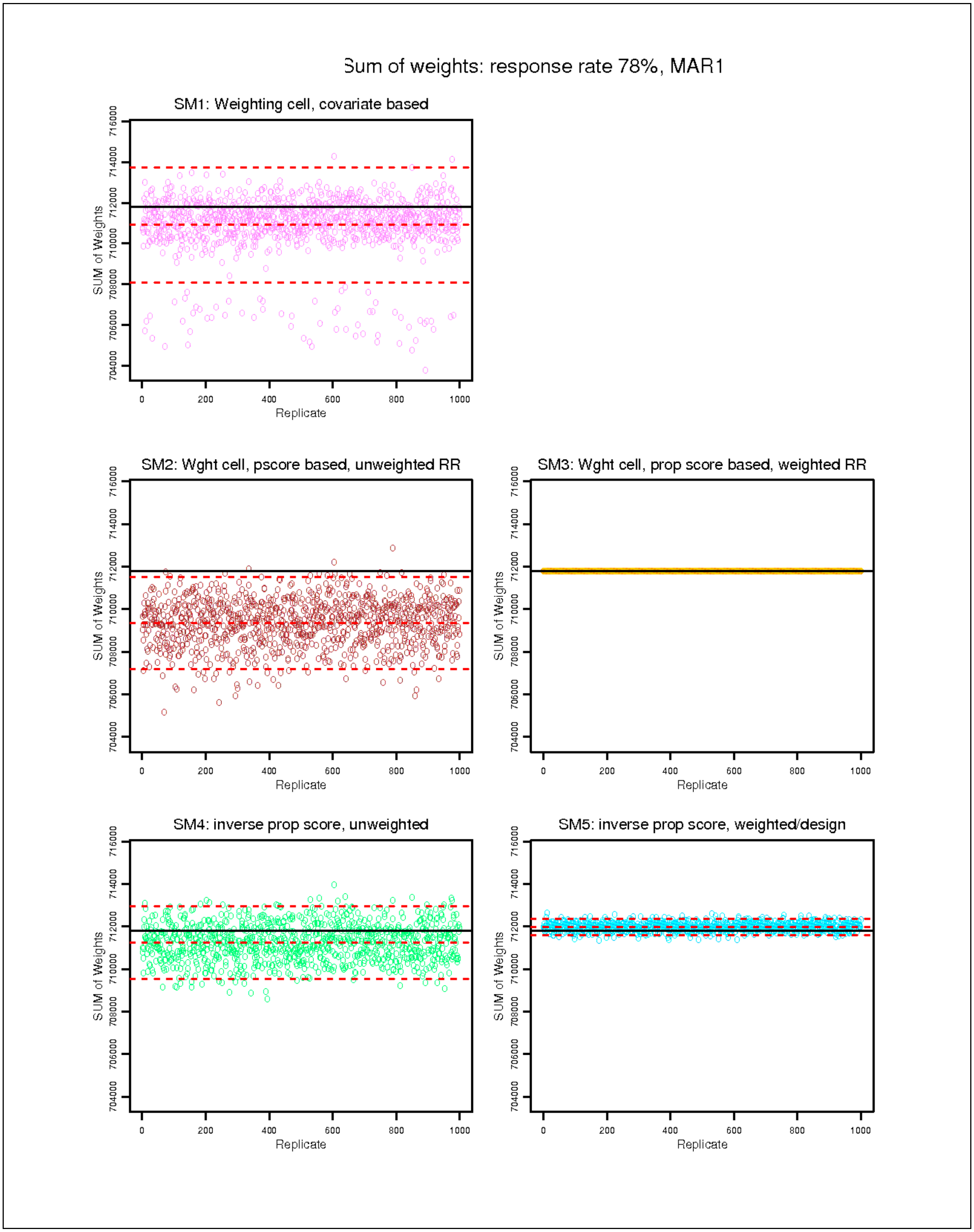


Figure 3: Over all total estimates plots under MAR1 with 78% response rate

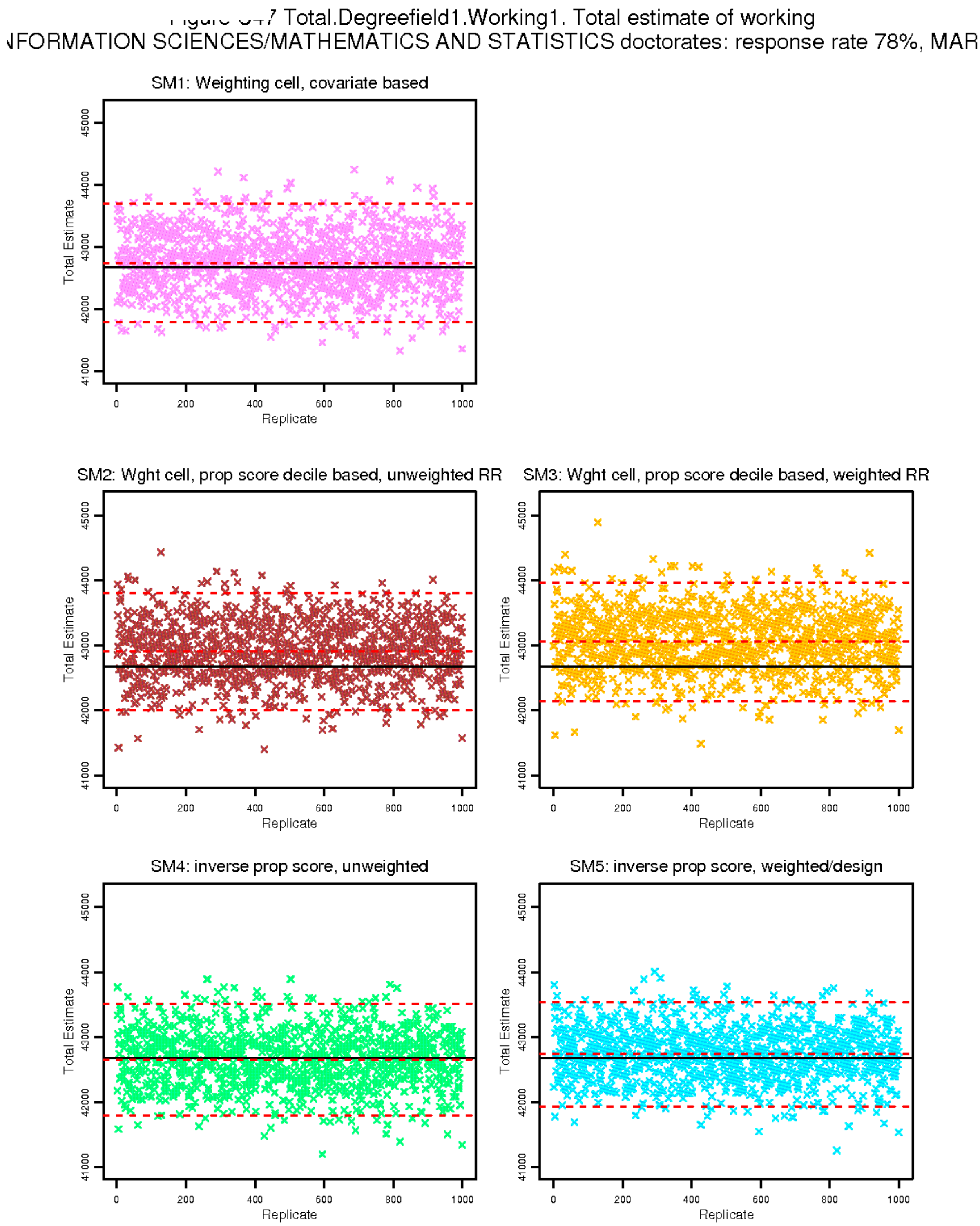


Figure 4: Domain total estimate plots under MAR1 with 78% response rate

Appendix

Table 3: Variables Included In SDR Weighting Research

Variable Name	**Description**	**Attributes**
GENDER	Sample member's gender	1=Male; 2=Female
ENTRYYR	Year of entry	1=1973, 1973-imputed, 1975, 1977, 1979 2=1981, 1983, 1985, 1987, 1989 3=1991, 1993, 1995, 1997, 1999 4=2001, 2003, 2006
SAMPTYPE06	2006 sample type	1=2003 Refuser; 2= 2003 Cooperative 3= 2003 NIR; 4=New Cohort;
CURCIT	Current citizenship	1=Citizen; 2=Noncitizen
HCAPIN06	Handicap status indicator for Sample Member in 2006	1=Yes; 2=No
BIRCIT	Citizenship at birth	1=Citizen at birth; 2=Non-citizen at birth
PHDFY	Ph.D. year	1=1960s and earlier; 2=1970s; 3=1980s; 4=1990s; 5=2000s
MARITAL	Marriage Status	1=Missing; 2=Married or living in a marriage-like relationship; 3=All others
DEPENDS	Number of dependents	1=Missing; 2=No dependent; 3=1 dependent; 4=2 dependents; 5=3 dependents; 6=4 or more dependents
PDWK1ED	Edited primary work activity	1=Missing; 2=Other; 3=Prof services; 4=Administration; 5=Teaching; 6=R & D
EDMOTHER	Mother's education level	1=Missing; 2=No bachelor degree; 3=Bachelor degree; 4=Master degree; 5=Ph.D. or Professional degree
EDFATHER	Father's education level	1=Missing; 2=No bachelor degree; 3=Bachelor degree; 4=Master degree; 5=Ph.D. or Professional degree
PHDCARN	Doctorate institution Carnegie Class	1=Doctoral/Research Universities-Extensive; 2=Doctoral/Research Universities-Intensive; 3=All others
RACE	Race/Ethnicity	1=Hispanic & all others; 2=White; 3=Black; 4=Asian
DEBTLEVL	In debt	1=Missing; 2=Yes; 3=No
PDOCSTAT	Employment status	1=Missing; 2=Not employed; 3=Employed
PHDCARNP	Doctorate institution Public\Private Indicator	1=Public institution; 2=Others
AGE40106	Age	1=Missing; 2=39 and under; 3=40 – 59; 4=60 and older
NSDRMED	SDR PhD Field of study (SESTAT Code Frame)	1=D's; 2=600's; 3=700's; 4=800's
DEGREEFIELD	SDR degree field	1=Information Sciences/Mathematics and Statistics; 2=Biological and Agricultural Sciences; 3=Health Sciences; 4=Physical and Related Sciences; 5=Social Sciences; 6=Psychology; 7=Engineering
LOCATED_03	2003 SDR location flag	1=Located in 2003 SDR; 2=Not located in 2003 SDR
ELIGIBILITY_03	2003 SDR eligibility known flag	1=Eligibility known in 2003 SDR 2=Eligibility unknown in 2003 SDR
RESPONSE_03	2003 SDR response flag	1=Responded to 2003 SDR 2=Did not response to 2003 SDR

Table 4: Average Bias for Median Salary Estimates by Degree Field and Gender

	Variable	Category	SM1	SM2	SM3	SM4	SM5
MCAR	Degreefield	0	-2	-1	-2	-2	-2
		1	265	265	264	264	264
		2	-271	-271	-272	-271	-271
		3	281	281	280	281	281
		4	85	86	86	85	86
		5	-87	-87	-87	-87	-88
		6	-48	-48	-48	-48	-48
		7	-18	-19	-19	-19	-19
	Gender	1	0	0	0	0	0
		2	-51	-52	-51	-52	-51
MAR1	Degreefield	0	48	-726	-850	61	51
		1	284	-341	-387	200	194
		2	-147	-1355	-1512	-70	-84
		3	446	8	-15	224	227
		4	117	-29	-34	126	124
		5	-139	-590	-640	-301	-308
		6	-43	-110	-119	-42	-47
		7	-46	-142	-169	-35	-35
	Gender	1	1	-16	-20	3	3
		2	58	-751	-840	8	-18
MAR2	Degreefield	0	38	-743	-796	2	2
		1	285	-185	-204	299	295
		2	-174	-1310	-1370	-218	-234
		3	514	148	135	445	444
		4	144	-14	-17	138	135
		5	-257	-608	-637	-233	-243
		6	-23	-130	-133	-65	-66
		7	-45	-158	-172	-44	-45
	Gender	1	0	-18	-20	-1	-1
		2	125	-626	-661	35	38
MAR3	Degreefield	0	34	-830	-901	-3	6
		1	219	-183	-204	221	238
		2	-229	-1428	-1491	-280	-261
		3	400	102	87	339	355
		4	111	-31	-33	111	112
		5	-212	-592	-631	-201	-185
		6	-5	-115	-119	-22	-21
		7	-49	-178	-194	-50	-49
	Gender	1	0	-26	-29	-1	-1
		2	60	-739	-781	-12	5
NMAR	Degreefield	0	-1981	-1982	-1982	-1981	-1981
		1	-3854	-4002	-3991	-3980	-3958
		2	-3278	-3459	-3450	-3301	-3294
		3	-558	-470	-472	-566	-538
		4	-3893	-4551	-4523	-4038	-4057
		5	-2637	-2668	-2665	-2585	-2582

		6	-738	-989	-970	-718	-705
		7	-4212	-4607	-4571	-4389	-4366
	Gender	1	-4883	-4984	-4980	-4914	-4920
		2	-1912	-1975	-1973	-1923	-1917

Table 3: Design Effects for the Weights

Design Effects for the Weights							
		Weighting Method					
Missing Mechanism	Response Rate	SM1	SM2	SM3	SM4	SM5	Full Sample Weight
MCAR	73%	1.114	1.114	1.114	1.114	1.114	1.114
MCAR	78%	1.114	1.114	1.114	1.114	1.114	1.114
MCAR	83%	1.114	1.114	1.114	1.114	1.114	1.114
MAR1	73%	1.203	1.150	1.156	1.198	1.199	1.114
MAR1	78%	1.197	1.150	1.155	1.200	1.197	1.114
MAR1	83%	1.149	1.150	1.155	1.211	1.148	1.114
MAR2	73%	1.201	1.159	1.162	1.195	1.196	1.114
MAR2	78%	1.198	1.159	1.162	1.198	1.200	1.114
MAR2	83%	1.196	1.158	1.161	1.214	1.220	1.114
MAR3	73%	1.213	1.153	1.157	1.203	1.202	1.114
MAR3	78%	1.206	1.153	1.157	1.204	1.199	1.114
MAR3	83%	1.204	1.153	1.157	1.213	1.206	1.114
NMAR	73%	1.224	1.169	1.168	1.268	1.248	1.114
NMAR	78%	1.212	1.153	1.156	1.235	1.220	1.114
NMAR	83%	1.206	1.152	1.155	1.239	1.218	1.114

Reference

Brick, JM, Kalton G. 1996: "Handling missing data in survey research". Statistical Methods in Medical Research, vol. 5, number 3, September 1996, pp. 215-238.

Kalton, G., and D. Kasprzyk, 1986: "The treatment of missing survey data." Survey Methodology, vol. 12, no. 1, 1986, pp. 1-16.

Little, R.J., and Vartivarian S. 2003: "On weighting the rates in nonresponse weights." *Statistics in Medicine*, vol. 22, pp. 1589-1599.

Sukasih, A., Jang, D., Vartivarian, S., Xu, H., Cohen, S., and Zhang, F. 2009: "Comparisons of Nonresponse Weighting adjustment Methods through Empirical Investigation and Simulation in Two SESTAT Surveys: National Survey of Recent College Graduates and Survey of Doctorate Recipients". Report submitted to National Science Foundation, upcoming.

Yang, M. 2007: "The 2006 Survey of Doctorate Recipients Weighting Implementation," Report Submitted to National Science Foundation, May 21, 2007.

Yang, M., Cox, B.G., Grigorian, K., and Sederstrom, S. 2007: "Sample Design and Implementation for the 2006 Survey of Doctorate Recipients". Report submitted to National Science Foundation, March 7, 2007.