NONRESPONSE ADJUSTMENTS FOR A SURVEY OF CHILDREN WITH DISABILITIES USING INFORMATION OF A RESPONSIBLE ADULT

Núria Díaz-Tena and Frank Potter
Mathematica Policy Research, P.O. Box 2393, Princeton, New Jersey 08543-2393

KEY WORDS: **Nonresponse, Propensity scores, Logistic Models, Supplemental Security Income, SSA Data Files**

1. Introduction

The National Survey of SSI Children and Families (NSCF) collected data on children with disabilities and their families who received or applied for Supplemental Security Income (SSI). The survey, sponsored by the Office of Research, Evaluation, and Statistics of the Social Security Administration (SSA), had two major objectives:

1. To provide information on the characteristics, experiences, and needs of the current cross-section of SSI child recipients and their families

2. To evaluate the effects of the Personal Responsibility and Work Opportunity Act of 1996 (P.L. 104-193; otherwise known as the Welfare Reform Act) on SSI children

The NSCF, which was administered in 2001-2002, was the first national survey of SSI children in more than 20 years. The survey was intended to fill a critical data need by providing current information on the health and well-being of SSI children and their families.

Mathematica Policy Research, Inc. (MPR) developed the sampling design for the NSCF questionnaire under a separate contract with SSA (Potter et al. 2000); implemented the survey (including data collection); created the coded variables, developed weights and variance estimation parameters for the survey; and imputed missing values for the income- coded variables.

A complex multivariate allocation algorithm was used to minimize survey costs and sample size subject to precision constraints for more than 150 survey estimates. The NSCF used a two-stage sample design. SSA administrative records were used as a list frame from which 75 primary sampling units (PSUs) were formed; after PSUs were selected, the list was used as the sampling frame for the selection of individual children within PSUs. The final sample size was 11,971 cases.

The NSCF sample design involves stratification and unequal probabilities of selection. Variance estimates calculated from NSCF data must incorporate the sample design features in order to obtain the correct estimate. For the NSCF, 75 PSU selections were made using a probability minimal replacement sequential selection procedure (Chromy 1979) and a composite size measure (Folsom, Potter, et al. 1987) accounting for eight strata of children in each PSU. The size measure was based on the number of children who had received or applied for SSI benefits. Within each PSU, children were stratified into eight sampling strata that classified the children on their SSI status in December 1996, whether they were subject to redetermination, and were continued or denied, their age, and their SSI status at time of selection (November 2000). We selected a single random sample in each stratum.

The population of applicants to SSI benefits at time of selection consisted of 4,374,545 records, 59 percent of them were applicants who were 16 years old or younger. These applicants had disabilities, and they sometimes could not answer the interview themselves. The adult who lived with or took care of the applicant was the person who answered the interview, or the person who refused to answer it.

Logistic propensity models for nonresponse adjustments have been used for various studies (Rosenbaum and Rubin 1983; and Kalton and Kasprzyk 1986). The logistic models estimate the propensity to respond based on attributes of both respondents and nonrespondents. In the current paper, we look at the advantages of using the personal characteristics of the adult responsible for the SSI applicant, in addition to the personal characteristics of the applicant, demographics, and previous SSI benefits received.

For the NSCF survey, we used weighted logistic models to adjust for nonresponse. There are two main causes of nonresponse: (1) when the child or knowledgeable guardian could not be located, and (2) when the child or the adult responsible of the child refused to complete the interview. For each cause of nonresponse, we first examined the pattern of nonresponse relative to the data available on sampled applicants. Weighted logistic regression models were used to estimate the location and participation propensity scores, which were used to adjust the initial weights for nonresponse. We introduced information about the adults responsible for the applicants in our logistic models to adjust for nonresponse, given that the adults who took care of the SSI applicants are the ones who could not be located or who refused to answer the interview. In addition, we tried to clarify the definition of the adult who is responsible for the applicant, using the data provided by SSA.

In Section 2, we describe in more detail how the weights were adjusted by the propensity scores. In Section 3, we compare the different ways in which to define the person who is responsible for the applicant, based on the data provided by SSA. Section 4 takes note of the improvement in the location logistic regressions if the variables describing the adults are introduced in the model. And in Section 5, we summarize which of the models better explains the nonresponse patterns, and the relationship between the person responsible for the applicant and the representative payee.

2. Nonresponse Adjustments for the NSCF

MPR released 11,971 cases to represent the SSI population of applicants since 1992. However, not all cases were interviewed see Table 1. Some cases could not be located (18.1 percent) and were classified as: unlocated by office, unlocated by field, or maximum number of calls with no contact. These cases represented 793,542 cases of the population of SSI applicants since 1992. The nonresponse adjustments for location distributed the unlocatable weights among the located cases. Some of the located cases did not participate in the survey (7.5 percent), and their weight was adjusted and distributed among the participating cases.

Table 1. Classification of Outcomes of the NSCF for Weighting Purposes

Classification of Outcomes	Cases Attempted	Weighted Percent
Total	**11,971**	100.00
Respondent	9,243	74.36
Eligible Complete	8,827	70.51
Ineligible Complete	516	3.98
Located Nonrespondent	782	7.50
Unlocatable	1,946	18.14

When the sample was drawn, an initial weight, W_I, was given to each selected case. This weight was the inverse of the probability of selection for the applicant. A weighted logistic regression with this initial weight was used to compute the probability of locating an applicant (propensity score for location). The propensity score for location, $P(L=1|x_i)$, is computed as in (1), where *L* stands for *Locate*, x_i, are a set of indicator variables that describe the applicant's characteristics (age, gender, type of disability, race), demographic information about the applicant (region, urban, move in the past year before the interview), SSI payments records (payment amount, years receiving SSI benefits, SSI status at time of survey, SSI status at time of Welfare Reform), and others. In equation (1), the *a* is the intercept of the logistic regression and the *b*'s are the coefficients for each one of the characteristics in the model. The inverse of the location propensity score is the location adjustment for the located cases.

The weight adjusted for location, W_L, is computed as in (2); it is the product of the initial weight and the inverse of the location propensity score. We then used another weighted logistic model (using the weight adjusted for location) to predict the probability that a located applicant would respond (propensity score for participation). The propensity score for participation, $P(Pa=1|L=1,x_i)$, is computed as in equation (3), where *Pa* denotes *Participation*. The inverse of the participation propensity score resulting from the application of this model can then be used as the participation adjustment, and (4) shows how to compute the final weight adjusted for participation, W_{Pa}, as the weight adjusted for location times the inverse of the participation propensity score:

$$P(L=1 \mid x_i) = \frac{e^{a+\sum_i b_i x_i}}{1+e^{a+\sum_i b_i x_i}} \tag{1}$$

$$W_L = W_I \times P(L=1 \mid x_i)^{-1} \tag{2}$$

$$P(Pa=1 \mid L=1, x_i) = \frac{e^{c+\sum_i d_i x_i}}{1+e^{c+\sum_i d_i x_i}} \tag{3}$$

$$W_{Pa} = W_L \times P(Pa=1 \mid L=1, x_i)^{-1} \tag{4}$$

Logistic regression models can use a large number of indicators to describe the nonresponse pattern for each applicant. In addition, logistic regression automatically distributes the weight of the nonrespondents among all the respondents. The inverse of the propensity score will distribute more weight to a respondent who has more similarities with the nonrespondents than to a respondent who is different to the nonrespondents. In addition, the propensity score approach is expected to reduce the potential for nonresponse bias (Diaz-Tena and Potter et al. 2002).

3. Adult Responsible for the SSI Applicant

The SSA data files contain extensive information about the characteristics of the SSI applicant and the benefits received from SSI. Information is also provided on the characteristics of the adult living with the child (applicant), the representative payee, and the living arrangements of the applicant.

The representative payee has categories for: mother, father, spouse, grandparents, child, relative, agency, and other ("Other" means that it has no representative payee, or the representative payee is not a relative).

The living conditions of the applicant describe whether the disabled child lives with parents, mother, father, spouse, or alone.

SSA updates the file when any of the child's characteristics change, and when the family structure changes. We will refer to the adult for whom the file was last updated as the relative whose records are in SSA's latest update. The relative's last update of the file can be for: both parents, the mother, the father, the spouse, or missing, and the most recent update of the file that provided information on the relative's age, gender, and disability status.

Table 2 shows the different percentages of representative payees (in the first row). The majority of representative payees are mothers (44.8 percent), and other (41.2 percent); there also are small percentages for relatives (7 percent, which includes the grandparents, child, and other relatives); fathers (3.9 percent); and agencies (3.1 percent). The first column of Table 2 shows the percentages of the adults for whom SSA has the last updated information in the files (denoted by relative). The largest percentages of relatives are mothers (52.7 percent), then both parents (29.7 percent), and, finally, small percentages for fathers and None.

Is the representative payee (the person who receives the check from SSA) the same person as the relative? We will answer this question by comparing which representative payees' records agree with the records for relatives. When the representative payee is the mother, 58.8 percent of the time the relative is the mother; and 37.2 percent of the time, the SSA files contain information on both parents. In other words, in 55.5 percent of the cases, the representative payee is the same person as the relative. This percentage was computed by multiplying the conditional probabilities of different type of relatives with the probabilities of the representative payee —that is, (1) when the representative payee is the mother and the relative is the mother or both parents; (2) when the representative payee is the father, and the relative is the father or both parents; (3) when the representative payee is the spouse and the relative is the spouse; (4) when the representative payee is the agency and there is no information of any relative; and (5), when the representative payee is Other, and there is no information on any relative.

Table 3 shows the percentages for the living condition of the applicant and the applicant's relationship to the relative. The largest percentages of living conditions of the disabled applicants are (shown in the first row): living with the mother (49.4 percent), living with both parents (28.2 percent), and disabled individual (18.9 percent). These two variables agree for 93.1 percent of the applicants; when (1) the applicant lives with both parents and we have information on both parents, (2) the applicant lives with the father and the relative is the father, (3) the relative is a disabled individual and we have no information of a relative, (4) the applicant lives with the mother and the mother is the relative, and (5) the applicant lives with the spouse and the relative is the spouse.

There is greater agreement between the living conditions of the applicant and the relative than in the relationship between the representative payee and the relative. That is, 44 percent of the relatives who take care of the applicants may not be living in the same household as the representative payee. On one hand, we have personal information on the relatives of the applicant, while, on the other, the SSA record provided the representative payee's address to locate the applicant.

4. Adults' Characteristic Relationship to Location Patterns

The two main causes of nonresponse are: (1) when the applicant was not located, and (2) when the applicant or the adult responsible for the applicant did not participate in the study. In this paper, we concentrate on the locating adjustment; the participation adjustment follows the same methodology. Table 4 gives the location rate (81.9 percent) for different characteristics of the relatives and representative payees. We want to study whether there is any pattern that shows a difference among the relatives and/or the representative payees, and the survey's ability to locate them. The characteristics of the relative are unknown when the SSA files had no information about the relative, or when there are missing values for the relative's characteristics. Adults older than 40, and the missing age cases are easier to locate (84 percent) than are younger adults (81 percent). Disabled adults have a lower location rate (80 percent) than those who are not disabled (82 percent). It is harder to find females (81 percent) than males (83 percent). In addition, an applicant whose representative payee is a grandparent is easier to locate (91 percent) than the rest (with percentages ranging from 62 to 86 percent).

We computed four different models for estimating the location propensity scores: (1) with characteristics of the applicant (age, race), demographic (region, moved in the last year), and economic information of the applicant (currently an SSI recipient, amount of money paid by SSI, years receiving SSI benefits); (2) same characteristics as the previous model, but also including characteristics of the relative (gender and disability status); (3) same characteristics as the first model, but also including

Table 4. Weighted Location Rates by Characteristics of the Relative and the Representative Payee

Characteristics	Weighted Percentage located
Age of the relative	
18-30 years old	81.4
31-40 years old	80.6
Over 40	83.3
Missing	84.6
Disability of the Relative	
Not disabled or missing	82.0
Disabled	80.0
Gender of the relative	
Female	81.1
Male	83.4
Missing	84.3
Representative payee	
Agency	82.2
Father	86.1
Grandparents	91.3
Mother	84.2
Other	78.1
Relative	82.6
Spouse	61.8

the relationship between the representative payee and the applicant; and (4) information about the applicant, the relative, and the representative payee's relationship to applicant.

The models were computed using SUDAAN to ensure a proper estimation of the standard errors. In addition to the main effects in the four models, we used second-order interactions. Even if we had information about the race of the relative, this information was not used in the logistic regression given the high correlation with the race of the applicant and the race of the relative. Because there are very few applicants whose representative payee is the spouse, this group will be combined with the relatives group (the representative payee who are defined as relatives) when computing the propensity score for location.

Table 5 shows the maximized R^2 adjusted, and the Akaike's information criterion (AIC) (Akaike, 1973) for the four described models. The first model uses only information about the applicant and has the lowest maximized R^2 and the largest AIC of the four models. Models 2 and 3 include more information than does model 1, and its maximized R^2 is larger than the maximized R^2 of the first model. Model 2 uses additional information about the relative in the model, and model 3 introduces the representative payee relationship in the model. Furthermore, these models obtained a lower AIC than model 1. When comparing model 2 and model 3, model 3 results in a higher maximized R^2 and a lower AIC than model 2. This means that introducing the representative payee information in the model improved the model more than when the information of the relative was added. Model 4 improves model 3 by obtaining a slightly larger maximized R^2 and a lower AIC than model 3. Model 4 provides more information than models 3 (besides the characteristics of the applicant and the representative payee, it also provides information of the relative)

Table 5. Significance of the four different location models

Model	Max-R^2 Adjusted	AIC
Applicant	0.0697	10,897
Applicant and relative	0.0810	10,839
Applicant and representative payee	0.0872	10,839
Applicant, relative, and representative payee	0.0936	10,794

Use of the maximized R^2 (Nagelkerke 1991) allowed comparison of the different models, given that each R^2 had been adjusted to achieve a maximum value of 1.

5. Conclusions

Logistic regression modeling is a useful method to adjust for nonresponse for list frame sampling design, especially when substantial data exist for the sample member.

For surveys collecting information on children or disabled individuals, it is useful to identify who is living or taking care of those individuals. The caretakers of family members are the persons who know the most about the disabled child and typically designated as the respondent. The caretakers must be found and interviewed to obtain the relevant survey information

The SSA files for SSI applicants provide very detailed information on the living conditions of the applicants, the relatives and the representative payee. Only 56 percent of the time, the representative payee is the relative for the SSI applicant population

In the case of location, the model that better estimates the location propensity scores is the model with information on the applicant, his or her demographics, SSI payment history, and the characteristics of the relatives and the representative payee. In our case, SSA provided the address of the representative payee, so it was sensible to introduce this information in the model; besides, the characteristics of the relative complemented the additional information about the person who care for the applicant and who answers the survey interview questions.

A recent article by Liao and McGee (2003) studied a new R2.adj that estimates an unbiased R^2 for logistic regressions. This seems to be a useful new measure for estimating the variation explained by a logistic model to the previous four models.

6. References

Akaike, H. "Information Theory and an Extension of the Maximum Likelihood Principle." *Second International Symposium on Information Theory.* Akademia Kiado, 1973, pp. 267-81.

Chromy, J.R. "Sequential Sample Selection Methods." *Proceedings of the Section on Survey Research Methods,* Alexandria, VA: American Statistical Association, 1979, pp. 401-406.

Diaz-Tena, N., F. Potter, M. Sinclair, and S. Williams. "Logistic Propensity Models to Adjust for Nonresponse in Physical Surveys." *Proceedings of the Section on Survey Research Methods* [CD-ROM]. Alexandria, VA: American Statistical Association, 2002.

Folsom, R.E., F.J. Potter, and S.R. Williams. "Notes on a Composite Size Measure for Self-Weighting in Multiple Domains." *Proceedings of the Section on Survey Research Methods,* Alexandria, VA: American Statistical Association, 1987; pp. 792-796.

Kalton, G., and D. Kasprzyk. "The Treatment of Missing Survey Data." *Survey Methodology,* vol. 12, 1986, pp. 1-6.

Liao, J.G., and D. McGee. "Adjusted Coefficients of Determination for Logistic Regression." *American Statistician*, vol. 57, no. 3, August 2003.

Nagelkerke, N.J.D. "A Note on a General Definition of the Coefficient of Determination." *Biometrica*, Vol. 78, 1991, pp. 691-768.

Rosenbaum, P.R., and D.B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrica*, Vol. 70, 1983, pp. 41-55.

Table 2. Percentages of the Representative Payee's Types, Relatives, and Conditional Percentages of the Relatives Given a Specific Representative Payee

		Representative Payee					
		Mother Pct. Weighted	Father Pct. Weighted	Spouse Pct. Weighted	Relative Pct. Weighted	Agency Pct. Weighted	Other Pct. Weighted
Relative	**Pct Weighted**	**44.78**	**3.90**	**0.04**	**6.96**	**3.10**	**41.22**
Both parents	**29.74**	37.17	56.90	4.64	11.62	15.95	23.20
Mother	**52.66**	58.81	4.33	37.35	38.92	32.76	54.37
Father	**3.45**	1.39	33.72		3.42	3.07	2.85
Spouse	**0.85**	0.35	0.11	51.18	0.42	0.17	1.54
None	**3.32**	2.28	4.93	6.75	45.62	48.05	18.03

Table 3. Percentages of Living Condition Categories, the Relatives and the Conditional Percentages of the Relatives Given a Specific Living Condition

	Living Conditions of the disabled Applicant				
Relative	Both Parents Pct. Weighted	Father Pct. Weighted	None Pct Weighted	Mother Pct. Weighted	Spouse Pct. Weighted
Weighted Pct	**28.22**	**2.69**	**18.90**	**49.42**	**.78**
Both parents	97.97		10.91	0.06	92.86
Mother	0.37		16.72	99.94	0.14
Father	1.65	100.00	1.55		2.44
Spouse			0.45		4.45
None			70.37		0.11