LOGISTIC PROPENSITY MODELS TO ADJUST FOR NONRESPONSE IN PHYSICIAN SURVEYS

Nuria Diaz-Tena, Frank Potter, Michael Sinclair and Stephen Williams
Mathematica Policy Research, Inc., Princeton, New Jersey 08543-2393

Key Words
Nonresponse; weighting; propensity modeling; weighting classes; Community Tracking Study; physician surveys

1. Introduction

Logistic propensity models for nonresponse adjustments have been used for various studies (Little 1986). The logistic models used to compute the scores reflect the propensity to respond based on attributes of both respondents and nonrespondents. Propensity scores can be used to compute explicit adjustments factors or to form weighting cells. Recent work has shown the benefits of using a limited number of weighting classes (Eltinge et al. 1997). In the current paper, we look at the advantages of using a different number of weighting cells or the propensity scores to adjust for nonresponse in a Physician Survey.

The Community Tracking Study (CTS), which is funded by the Robert Wood Johnson Foundation, is designed to provide a sound information base for decision making by health leaders. It does so by collecting information on the United States health system, and how it is evolving, as well the effects of those changes on people. Begun in 1996, the CTS, is a longitudinal project that relies on periodic site visits and surveys of households, physicians, and employers. This survey consists of two samples, a site sample and a supplemental sample. The site sample is a national survey of 60 locations in the United States: 48 large Metropolitan Statistical Areas (MSAs), 3 small MSAs, and 9 non-MSAs. The supplemental sample includes all 48 contiguous states stratified in 10 different regions, as described in Potter et al. (2000).

In the Physician Survey, we had three different subgroups of physicians for the site sample, and for the supplemental sample based on their Round Two interview status: (1) Round Two interviews (reinterviews); physicians who completed the Round Two interview, (2) Round Two noninterviews (noninterviews); physicians who were selected for the Round Two sample but who did not complete the interview for reasons such as ineligible, refusals or not located, and (3) new sample (new); physicians in the Round Three sampling frame who were not selected for the Round Two sample.

For the physician survey of the Round Three, we used weighted logistic models to adjust for nonresponse. There are two main causes of nonresponding: (1) when the physician could not be located, and (2) when the physician refused to complete the interview. For each cause of nonresponding, we first examined the pattern of nonresponse relative to the data available on sample members. Then, because of the different rates of nonlocation and nonresponse in each subgroup, we used different models in each subgroup and sample to adjust for nonresponse. Weighted logistic models were used to predict the probability of locating a physician (propensity score for location) for each of the three subgroups in the site and the supplemental samples. For each subgroup in the site and supplemental samples, we then used other weighted logistic models to predict the probability that a located physician would respond (propensity score for respondents). The inverse of the location and response propensities resulting from the application of those models -total inverse propensity score for nonresponse-, can then be used as the adjustment factor to the weights.

There are other possible methods for adjusting for nonresponse. The weighting classes or cells approach defines cells that have approximately equal response probabilities within the cells. The respondent weights are then inflated in each cell to account for nonrespondents. One way to define the cells is given in Little (1986), Eltinge et al. (1997), and Smith et al. (2001), sampling units are grouped according to the propensity scores obtained by logistic models.

In the present paper, we compare logistic propensity models and weighting classes, as Carlson et al. (2001) did. However, we constructed the weighting cells with propensity scores focusing on the bias, variance, and root mean square error of the logistic regression and weighting classes with different number of cells.

2. Response Rates for the Different Subgroups in Each Sample

The subgroups in both samples have different percentages for location and responce, as Table 1 shows. The peculiarities are captured by modeling each subgroup in each sample with a particular logistic regression model that better explains the location and response patterns in each case.

The location rate for the reinterview subgroup is the highest of all the groups: 98.5 percent for the site sample and 98.6 percent for the supplemental sample. Physicians in this subgroup were found at higher percentages because they were located and had completed the survey in Round Two. The location rate for noninterviews is 90.2 percent for the site sample, and 91.9 percent for the supplement sample. The location rate for the new group of physicians in Round Three is similar to that of the previous groups: 90.2 percent for the site sample, and 91.0 percent for the supplement sample, respectively. The logistic regression models for location were used to compute propensity scores for the physicians who were located. The location adjustment factor is the inverse of the location propensity score, which itself is the inverse of the predicted probability of locating a physician. The location adjustments factors compensated less than 10 percent of the sampled physicians. The response rates vary more among subgroups than the location rates. The response rates for reinterview physicians who are located are 81.2 percent for the site sample and 83.0 percent for the supplement sample. This subgroup of physicians collaborated and completed the interview in Round Two, as expected these are the physicians who obtained the highest response rate in Round Three. The response rates for the noninterview physicians who were located are only 38.5 percent for the site sample, and 38.2 percent for the supplement sample. This subgroup included: unlocated, refusals, and ineligible physicians of the Round Two Survey, and had the lowest response rates because the refusals in the last round are the hardest people to obtain a complete interview. The response rates of the new physician sample that were located are 60.0 percent for the site sample and 53.9 percent for the supplement sample. The response rates fall between the reinterviews and noninterviews responses rates. This subgroup contains the group of physicians with no past involvement with the survey and no previous records of their willingness to provide information about their practices..

The logistic regression models for respondents were used to compute propensity scores for physicians who completed the survey. The response adjustment factor is the inverse of the predicted probability of a physician completing a survey. In Round Three, the adjustments factors have to compensate for approximately 18 percent for the reinterviews, 62 percent for the noninterviews, and 45 percent for the new subgroup of the sampled physicians who were located but did not complete the interview.

3. Logistic Regression Models to Compensate for Nonlocatable and Nonresponding Sample Cases

To prepare the logistic models, we used a weighted forward stepwise logistic regression procedure from SAS to select variables. This procedure indicates the significance of main effects, second and third order interactions when they are introduced into the model. We obtained a full logistic regression model using the more significant main effects, second and third order interactions. Any combination of main effects and second order interactions involved in the third order interactions was included in the full model, regardless of their significance. Then we used this full model in SUDAAN, which computes the correct sampling variances for the estimates of the models and takes into account the sampling design of the survey, to eliminate the predictors that are not significant.

The variables used in the logistic regression models are: age, board certification, country of medical school, gender, income (in reinterviews), past disposition code (in noninterviews), present employment, region, specialty, and urban/rural status. Besides these variables, second and third order interactions were included if significant in the model.

The categories were chosen depending on the number of observations in each category and the different location or response rates in each category. For example, the categorization of specialty in the location model for the noninterviews, we used only four categories: General/Family Practice, Internal Medicine, Pediatrics, and Other Specialties. However, for the response model for the noninterviews we used six categories: General/Family Practice, Internal Medicine, Pediatrics, Surgeons, Psychologists and Other Specialties.

This paper shows the results of two models: (1) the location model for the new subgroup in the site sample, which adjusted for 9.8 percent of the physicians that could not be located, and (2) the response model for the new subgroup in the site sample, which must adjust for 40.0 percent of the located physicians that did not respond. We limited the analysis to these models because this would be more similar to one-time survey situation. The conclusions, however, are based on the six logistic regression models from the site sample.

The predictors for the location model for the new physician sample in the site sample are: age (less than 40, 40-49, and more than 50), board certification (certified, not certified), country of medical school (USA/Canada, other), urban/rural (urban and rural), present employment (solo or two-physician practice, group practice, and other), region (Northeast, Midwest, South, and West), specialty (Gen/Family Practice, Internal Medicine, Pediatrics and Other Specialties), and gender (male and female). The model includes five third order interactions, two second order interactions, besides the main effects and the second order interactions that are combinations of the third order interactions included in the model. Twenty-six variables are used in this model, with an R^2 of 0.046. Table 2A indicates the significance of the different main effects, second order interactions, and third order interactions for the location model for the new physicians in the site sample.

The predictors for the response model for the new physicians in site sample are: age (less than 40, 40-49, 50-59, and more than 59), board certification (certified, not certified), country of medical school (USA/Canada, other), urban/rural (urban and rural), present employment (solo or two-physician practice, group practice, and other), region (Northeast, Midwest, South, and West), specialty (Gen/Family Practice, Internal Medicine, Pediatrics, Surgeons, Psychologists, and Other Specialties), and gender (male and female). The model includes eight third order interactions, five second order interactions, besides the main effects and the second order interactions that are combinations of the third order interactions included in the model. There are 38 variables used in this model, with an R^2 of 0.068. Table 2B indicates the significance of the different main effects, second interactions and third interactions for the response model for the new physicians in the site sample. More details on the selection of the model are available in Diaz-Tena et al. (2002).

4. Weighting Class Method

The main idea is to define groups, or cells, of sample units that are believed to have approximately equal response probabilities. For example, the respondents with similar propensity score for location; including those who were located, and those respondents who were not located, are supposed to have a similar behavior toward locatability. The same situation applies to the located respondents with similar nonresponse propensity scores, which have a similar behavior toward completion of the interview. The cells are formed by direct grouping of sample units according to their estimated response probabilities. Then k cell divisions are defined by the estimated k-1 quantiles of the propensity scores. This equal-quantile method gives some control over the expected number of respondents in each cell. Principal attention is directed to the sensitivity of results to the number of cells used.

Within each cell, the respondents weights are increased to compensate for the nonrespondents, under the assumption that they are alike. To ensure relatively stable adjustments, some rules are usually used when forming the cells, such as ensuring a minimum number of respondents per cell of say 20. We did not maintain the other common rule of keeping the ratio of respondents to nonrespondents in the cell smaller than two. In the noninterviews subgroup of the site sample, there are only 35 percent of respondents, and, when using the weighting class method with only one class, its smallest adjustment factor was greater than 2.6.

5. Comparison of the Different Weighting Adjustments

We compared the sets of weights: the weighting class only approach with different number of cells ranging from 1 to 50, and the weighting class propensity model approach. We computed the relative bias, standard error and root mean square error for each age, board certification, country of medical school, present employment, region, and gender categories.

We computed the percentages for the population of physicians for each category of the variables from the sampling frame. For nearly all sampled physicians, demographic, personal, and practice characteristics are available from the American Medical Association (AMA), and the American Osteopathic Association (AOA) files that were used as the sample frame. We estimate the percentage of physicians in each category for the weighting class only approach for all of the different numbers of cells and the percentage of physicians in each category for the weighting class propensity model approach. The relative bias, standard error, and root mean square error are compared with each of the adjustments for nonresponse methods. The plots show the results for age, employment, and the average of the previously mentioned characteristics.

On the one hand, the plots show that the weighting class approach with only one cell has the largest bias, and that the bias decreases as more cells are used in the weighting class approach. The use of the actual propensity scores for nonresponse adjustments obtains the lowest bias. On the other hand, the standard error is larger when the actual propensity scores are used and it is smaller when weighting classes are used with fewer cells. We observe the classical bias/variance tradeoff. The root mean square error plot shows that the propensity score approach is equivalent to the weighting class approach (when forming cells with the propensity scores), when there are between eight and ten classes. The small number of classes reduces the standard error, and the large number of classes reduces the bias.

6. Conclusions

We found a large difference between the propensity score approach and the weighting cell method with very few cells when looking at the CTS Physicians Survey Round Three. The benefit of the propensity score approach is a reduction in bias, but this method has greater variation in the weights than does the weighting cell method. The benefit of the weighting cell method is a small variation in the weights as the number of cells decreases, but this method has more potential bias as the number of cells decreases compared to the propensity score method or when a larger number of cells in the weighting cell method. The minimum number of cells in the weighting cell method sufficient to decrease the bias and maintain a small variation of the weights is between eight and ten.

Overall, we feel that there is less effort required in the propensity method. Modeling is very labor intensive, but it has to be performed in either method. The weight adjustments for the propensity modeling are computed after obtaining the propensity scores of the best model. For the weighting cell method, however, the cells are formed after computing the propensity scores. Once the cells are formed the adjustments within cells can be computed. We chose propensity score method over the weighting class method with eight to ten cells for the easier implementation of the method.

7. Future Research

This paper has focused on a comparison of the propensity scores method with the weighting cell method when cells are formed according to the propensity scores of the logistic model before poststratification and trimming. The same comparison could be made after poststratification and trimming.

It is possible to form the cells with the more significant variables of the logistic model, that is, the variables with the highest odds ratio. This method of cell formation will group cases with similar nonresponse propensity scores and equal characteristics of the independent variables. That is a similar method used by Tambau et al.(1998), in which they used CHAID to form the weighting cells.

8. References

Carlson, B.L., and S. Williams. “A Comparison of Two Methods to Adjust Weights for Nonresponse: Propensity Modeling and Weighting Class Adjustments.” Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM]. Alexandria, VA: American Statistical Association, 2001.

Diaz-Tena, N., F. Potter, S. Williams, R. Strouse, and M. Ellrich. “Report on Survey Methods for the Community Tracking Study’s 2000-2001 Round Three Physicians Survey.” Forthcoming in Center of Studying Health Systems Change Web site www.hschange.com.

Eltinge, J.L., and I.S. Yansaneh. “Diagnostics for Formation of Nonresponse Adjustments Cells, with an Application to Income Nonresponse in the U.S. Expenditure Survey.” *Survey Methodology*, vol. 23 1997, pp. 33-40.

Little, R.J.A. “Survey Nonresponse Adjustments for Estimates of Means.” *International Statistical Review*, vol. 54 1986 pp. 139-157.

Potter, F., M. Sinclair, and S. William. “Examining Attrition in the Physicians Component of the Community Tracking Study.” Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM]. Alexandria, VA: American Statistical Association, 2000.

Smith, P.J., J.N.K. Rao, M.P. Battaglia, T.M. Ezzati-Rice, D. Daniels, and M. Khare. “Compensating for provider Nonresponse Using Response Propensity to Form Adjustment Cells: the National Immunization Survey.” *Vital and Health Statistics*, vol. 2 2001, pp.133.

Tambau, J.-L., I. Schiopu-Kratina, J.Mayda, D. Stukel, and S. Nadon. (1998). “Treatment of Nonresponse in Cycle Two of the National Population Health Survey.” *Survey Methodology*, vol 24 1998, pp. 147-156.

Table 1: Location, Response, and Response I for Located Physicians for the Different Strata and Surveys

Subgroup	Survey	Sampled	Sample Located	Sample Respondents	Weighted Location	Weighted Response Among Located	Weighted Response
Reinterviewed	Site	10,345	10,160	8,341	98.5%	81.2%	80.0%
	Supplement	1,049	1,035	858	98.6%	82.9%	81.8%
Noninterviewed	Site	6,682	5,999	2,332	91.1%	38.5%	35.1%
	Supplement	633	581	221	91.9%	38.2%	35.1%
New	Site	5,561	4,970	2,993	90.2%	60.0%	54.1%
	Supplement	670	606	321	91.0%	53.9%	49.0%

Table 2.A.[1]: Significance of the Location Model for the New Subgroup in the Site Survey

Main Effects	Age	Certified	Country	Urban	Employment	Region	Specialty	Gender
	-			-	*	*	**	
Second Order	Age	**	No	No	-	No	No	
Interactions		Certified		No	No	-	*	
			Country		No	*	*	
				Urban	No	No	No	No
					Employment	No	No	No
						Region	**	No
							Specialty	*
								Gender
Third Order	Age	Certified	Gender					
Interactions	County	Certified	Specialty	*				
	Country	Region	Specialty	**				
	Country	Specialty	Gender	*				
	Disposition	Employment	Gender	-				

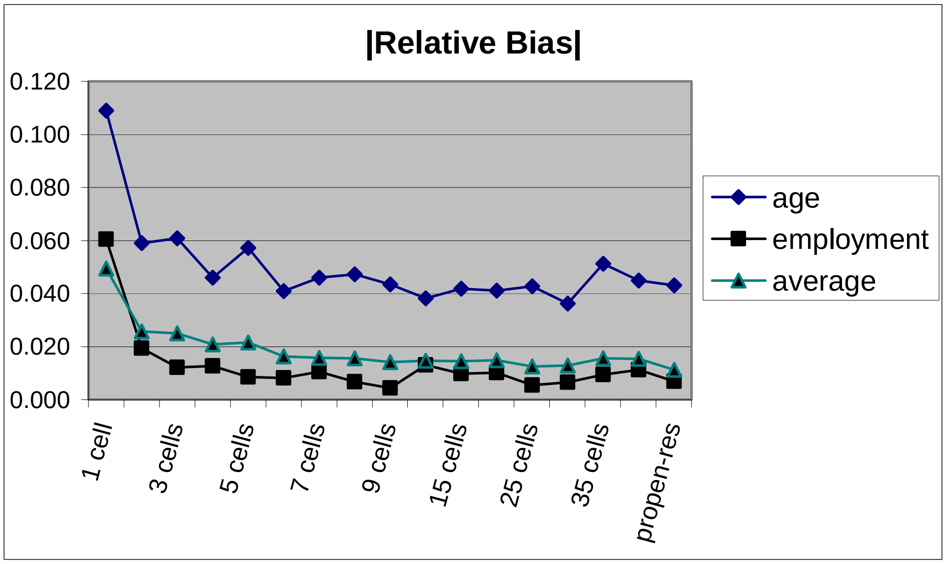
Table 2B: Significance of the Non-Response Model for the New Subgroupin the Site Survey

Main Effects	Age	Certified	Country	Urban	Employment	Region	Specialty	Gender
						*		
Second Order	Age	No	-	*	-			
Interactions		Cdrtified				*	**	No
			Country					
				Urban	*	-	No	
					Employment	*	No	
						Region	No	*
							Specialty	No
								Gender

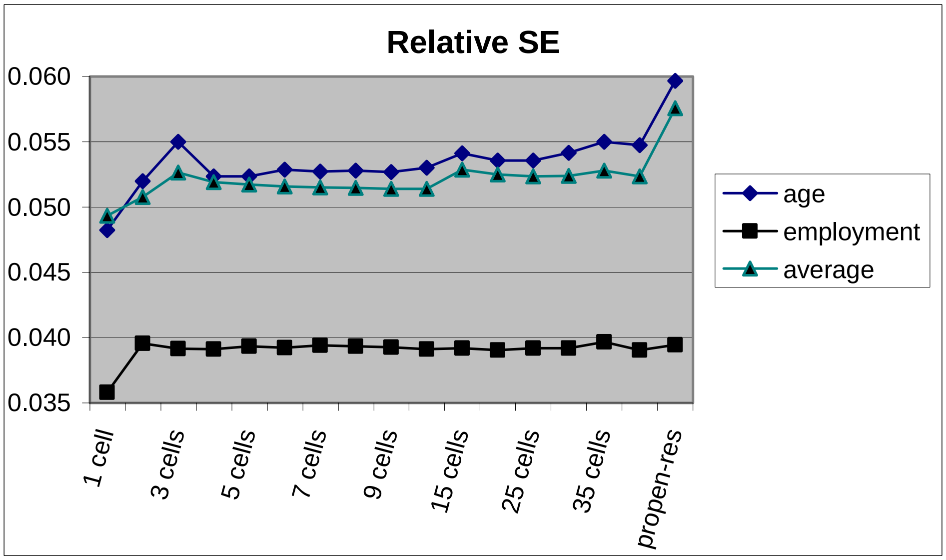
Third Order	Age	Country	Gender	*	Certified	Urban	Region	*
Interactions	Age	Employment	Gender	-	Certified	Urban	Employment	*
	Certified	Country	Region	-	Country	Urban	Region	**
	Urban	Region	Gender	-	Country	Employment	Region	*

[1]The significance levels are noted by: *** the smallest p value < 0.001, ** the smallest p value < 0.01, * the smallest p value for <0.05, - the smallest p value < 0.1, and blank denotes the smallest p value ≥ 0.1.

PLOT 1. Relative bias computed for different methodologies to adjust for nonresponse for different characteristics of the physicians for the new subgroup of the CTS Round Three Physician Survey.



PLOT 2. Relative standard error computed for different methodologies to adjust for nonresponse for different characteristics of the physicians for the new subgroup of the CTS Round Three Physician Survey.



PLOT 3. Relative root mean square error computed for different methodologies to adjust for nonresponse for different characteristics of the physicians for the new subgroup of the CTS Round Three Physician Survey