Modeling of Response Propensities in the Joint Canada/United States Survey of Health

Meena Khare and Karen E. Davis, National Center for Health Statistics,
3311 Toledo Road, Room 3218, Hyattsville, MD 20782

Key Words: Logistic Regression, Weighting class, Nonresponse, Noncoverage

Background

The Joint Canada/United States Survey of Health (JCUSH), a one-time national random-digit-dialing telephone survey, was conducted jointly by the National Center for Health Statistics (NCHS) and Statistics Canada from November 2002 to July 2003. The JCUSH was designed to compare selected health characteristics on both the American and Canadian adult population 18 years of age and older (Statistics Canada, JCUSH: Overview, 2004). Only one adult was selected from each responding household. Persons aged 65 years and older were sampled at a higher rate than other adults. For consistency, the same sampling and data collection methodologies were used in both countries. A total of 5,183 American and 3,505 Canadian adults participated in the JCUSH.

Weighting class adjustments are commonly used in government sponsored sample surveys to adjust for nonresponse. This method assumes that the responses of nonrespondents, if obtained, would be similar to those from respondents within homogenous weighting class cells. The variables used to create the homogeneous weighting class cells, should be available for both respondents and nonrespondents, however, to ensure adequate cell sizes ($\geq$ 20), only a limited number of demographic and geographic variables can be used to form the weighting class cells. Within each weighting class cell (e.g., Hispanic males aged 18-44 years living in Northeast), the inverse of the weighted response rate is used to adjust the sampling weights for the respondents in that cell. An alternative approach for forming adjustment cells is to use predicted response propensities from a logistic regression model (Ezzati-Rice, et al., 1994). To obtain predicted response propensities, a binary response indicator variable is regressed on potential demographic, socioeconomic and geographic covariates that are related to nonresponse, and are available for both respondents and nonrespondents. In telephone surveys, households without telephone service are excluded from the sample, and are nonrespondents. Therefore, weighting methods need to incorporate adjustments for the noncoverage of nontelephone households in telephone surveys.

Introduction

It is well known that telephone surveys are subject to coverage bias because of noncoverage of nontelephone households. The problem of nontelephone coverage bias may be viewed as attempting to define weighting classes that make the missing data ignorable rather than nonignorable, as is often the case (Ferraro, et al., 2001).

Approximately four percent of the U.S. households do not have a telephone inside their house (Table 1, Current Population Survey, 2002). Table 2 shows that the rates of telephone coverage show substantial variation by geographic, demographic and socioeconomic factors. In addition, elderly persons were oversampled in the JCUSH to meet the target sample size. Because only one adult was selected per household and an elderly person was selected with certainty, adults aged 18-64 years living with elderly (LWE) persons aged 65+ years were inadvertently excluded in the JCUSH sample. From the 2002 National Health Interview Survey (NHIS), we estimated that 4.5 percent of adults aged 18-64 years lived with an elderly person (Table 3) in the U.S. In this paper we investigate model-based response propensity methods to adjust for the noncoverage of nontelephone households and compare the weighted estimates with those obtained using the standard simple poststratification method.

Objectives

After adjusting the base sampling weights for households with multiple telephone lines, selection of one adult per household, and interview nonresponse, the next major task for the JCUSH weighting strategy was to develop two propensity models to adjust for noncoverage of nontelephone households and for adults aged 18-64 years living in households with elderly persons. Because NHIS collects data on the age of all members of the responding household and also includes all households with and without telephone service, we used data from the 2002 NHIS to develop the weighting methodology for the JCUSH/U.S.A. Our research included the use of 2002 NHIS data to compare selected characteristics

of adults living in telephone and nontelephone households, to compare characteristics of adults aged 18-64 years living in households with or without an elderly person aged 65 years and older in the household, and to assess the magnitude of the potential bias due to noncoverage.

Methodology to adjust for noncoverage of nontelephone households

Using data from the 2002 NHIS, we divided the household sample into telephone and nontelephone households based on the reported telephone status. *("Is there at least one telephone INSIDE your home that is currently working?"*) Table 2 compares the demographic and health characteristics of the two subgroups and shows the magnitude of the bias in terms of the difference in percentage points (B-A). From Table 2, nontelephone status appears to be related to age 18-44 years and socioeconomic status of the household members (e.g., having low income, being black or Hispanic, having no health insurance at all, or no private health insurance).

To develop a propensity model, we identified a set of variables that were common to the 2002 NHIS and the JCUSH (U.S.A. as well as Canada) and that were associated with nontelephone status. In order to evaluate the propensity model, data from the 2002 NHIS was split into a test sample and a model sample. The test sample was created by selecting a random sample of all adults living in telephone households yielding approximately 5,000 adults (similar to the JCUSH/U.S.A. sample). The model sample included the remaining 35,178 adults from the telephone and nontelephone households. Table 2 shows that characteristics of the adults in the test sample are similar to those of adults from the NHIS telephone sample. The model sample was used to fit a logistic regression model to predict telephone status. Since telephone status at the individual case level was dichotomous, a binomial logit model was used (Battaglia, *et al.*, 1995). The logit model was defined as

$$\ln\{\tau_i /(1-\tau_i)\} = x_i\beta + e \qquad (1)$$

where τ_i is the probability that the person i lives in a household that did not have telephone service. The beta coefficients from the final model including age group, sex, race/ethnicity, marital status, education, household income, house tenure, and Census region as covariates, was used to estimate propensity of telephone status in the test sample.

We then evaluated three methods to adjust for noncoverage of nontelephone households: simple poststratification (SP), direct use of propensity scores from the model (Method 1), and a weighting class method using quintiles of the propensity scores and demographic characteristics (Method 2). The simple poststratification method used the 2002 NHIS population control totals corresponding to the Census region, age group, sex, and race/ethnicity variables [based on the population controls from the 2002 Current Population Survey (CPS)]. Method 1 used the predicted propensity scores from the NHIS model as an adjustment factor [=1+exp(logit(τ_i)] to the nonresponse-adjusted NHIS base sampling weight in the test sample. This step adjusted the weight of a person in a telephone household to account for similar persons from nontelephone households. For Method 2, weighting class cells were formed based on quintiles of the propensity scores, age, and race/ethnicity. An adjustment factor was then computed in each cell equal to the weighted total population divided by the weighted total in the test sample of telephone households. This factor was then applied to the nonresponse-adjusted NHIS base sampling weight. Finally, we added a final poststratification adjustment to Methods 1 and 2 to adjust the total cell counts to the total U.S. population (based on the population controls from the 2002 CPS). The final post-stratified weights were then used to compute weighted estimates and to compare the three methods. To account for complex sample design of the NHIS, SUDAAN (Shah et el., 1997) was used to analyze the data.

Methodology to adjust for noncoverage of adults aged 18-64 years living with at least one elderly person

Using 2002 NHIS data, we created a binary variable as LWE=1, if an adult aged 18-64 years lived in a household of size $\geq$2 containing at least one elderly person of age 65 years or older, and LWE=0 for all others. Analysis was limited to adults aged 18-64 years because the noncoverage of LWE adults only affected this age group. We used demographic, socioeconomic, and health information to identify characteristics of the LWE adults and to evaluate the magnitude of the potential bias due to noncoverage. Table 3 compares the percentage of adults with LWE=1 in the telephone households, nontelephone households, and all NHIS households by selected characteristics. Overall, 4.7% of adults aged 18-64 years lived with at least one elderly person in telephone household and only 1.7% of adults living in nontelephone households lived with an elderly person 65 years or older. Table 3 shows that LWE adults are about three times more likely to live in a

telephone household than in a nontelephone household. It also shows that most of the LWE adults are minorities (i.e., Hispanic, non-Hispanic black, or non-Hispanic others), females, home owners, and with income ≤ $20,000 or >$20,000.

Because we did not have information on the LWE status in the JCUSH sample at the time of this research, instead of using a model-based method to reduce potential bias due to noncoverage of LWE adults in the JCUSH sample, a simple poststratification method was used. Further research is planned to investigate application of model-based methods to adjust for this noncoverage.

Results

To assess the impact of the three noncoverage adjustment methods, weighted estimates of several health and sociodemographic characteristics are compared. We used self-reported excellent/very good health status and fair/poor health status, no private health insurance, no health insurance at all, income ≤ $20,000, and education as college graduate for comparison. Estimates of bias were computed by taking the difference between the NHIS estimates and the estimates from Methods 1, 2 and SP. Standard errors were calculated using SUDAAN (Shah et al., 1997) and the ratio of the NHIS standard errors to the standard errors from Methods 1, 2 or SP were compared.

There appear to be substantial differences between the characteristics of the telephone and the nontelephone households when we compared the sociodemographic characteristics (Table 2). However, after applying adjustments for the noncoverage to sampling weights in the test sample, no differences were observed in the point estimates or the ratio of standard errors when using simple poststratification or weighting class Methods 1 and 2 (Table 4). Because no significant differences were found between estimates using Methods SP, 1 or 2, we selected the SP method to adjust for noncoverage of nontelephone households (Table 5). Table 5 shows a comparison of the weighted estimates from the 2002 NHIS and the test sample. SP method was also used in the JCUSH/Canada sample to adjust for the noncoverage of nontelephone households (see Public Use Microdata File User Guide, Statistics Canada, 2004).

Finally, after implementing the simple poststratification method to adjust for noncoverage of nontelephone households and LWE adults, we compared the final age-specific weighted estimates from of selected characteristics from the JCUSH/U.S.A. with the 2002 NHIS estimates. Table 6 shows the differences in weighted estimates by age groups. The largest difference in estimates (i.e., health status: excellent or very good) was among adults aged 18-44 years and most of the adults aged 18-44 years live in a household without telephone service.

Discussion

From the analysis, it is clear that the magnitude of noncoverage adjustment is smaller than the expected sampling error from the three methods. That is, any confidence intervals derived for the selected characteristics based on simple poststratified weights included the estimates based on the noncoverage-adjusted weights from alternative methods. This may not be surprising, given the relatively small sample size of the JCUSH.

The propensity model developed for the noncoverage adjustment was necessarily weak given the limitation of using only variables common to the NHIS and JCUSH. In fact, the propensity model was built primarily using the same geographic and demographic variables (age, sex, race/ethnicity, and Census region) that were incorporated into the SP method.

A simple poststratification method was used in the JCUSH/U.S.A. and the JCUSH/Canada samples to adjust for the noncoverage of nontelephone households and the noncoverage of LWE adults aged 18-64 years. The Canadian sample used Province, age group and gender to create poststratification cells whereas the American sample used the Census region, age group, sex, and race/ethnicity to form the poststratification cells. It was concluded that to ensure a sufficient representation of elderly persons aged 65 years and older in the JCUSH, a possible bias due to noncoverage of LWE adults aged 18-64 years was an acceptable compromise (Statistics Canada, JCUSH: Overview, section on sampling, 2004).

Finally, significant differences were observed between the final weighted estimates from the JCUSH/U.S.A. and the 2002 NHIS. Some of these differences could be attributable to sampling errors due to the fact that the JCUSH/U.S.A. had much lower response rate (50.2%) than the 2002 NHIS (89.6%).

References

Battaglia, M., Malec, D., Spencer, B., Hoaglin, D., and Sedransk, J. (1995) Adjusting for Noncoverage of Nontelephone Households in the National Immunization Survey, 1995 Proceedings of the Section on Survey Research Methods, Alexandria, VA: American Statistical Association.

Current Population Survey, October 2002: Basic Monthly Survey [machine-readable data file] conducted by the Bureau of the Census for the Bureau of Labor Statistics. - Washington: Bureau of the Census [producer and distributor], 2004.

Ezzati-Rice, T., and Khare, M. (1994) Modeling of Response Propensity in the Third National Health and Nutrition Examination Survey, 1994 Proceedings of the Section on Survey Research Methods, Alexandria, VA: American Statistical Association.

Ferraro, D., and Brick, J.M. (2001) Weighting for Nontelephone Households in RDD Surveys, 2001 Proceedings of the Section on Survey Research Methods [CD-ROM], Alexandria, VA: American Statistical Association.

National Center for Health Statistics (2003). Data File Documentation, National Health Interview Survey, 2002 (machine readable data file and documentation). National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland.

Shah, B.V., Barnwell, B.G., and Bieler, G.S. (1997), SUDAAN User's Manual, Release 7.5., Research Triangle Institute, Research Triangle Park, NC

Statistics Canada, (2004). Joint Canada/United States Survey of Health (JCUSH): Overview, http://www.statcan.ca/english/concepts/jcuhs/

Statistics Canada, (2004). Joint Canada/United States Survey of Health (JCUSH): Public Use Microdata File User Guide. http://www.statcan.ca/english/sdds/document/5020_D3_T9_V1_E.pdf

Appendix

Table 1. Rates of telephone coverage in the U.S.A. by age group, Current Population Survey (November, 2002)

Age Group	Phone Inside Household (%)	Phone Not Inside Household (%)
Total	96.12	3.88
18-44 years	95.08	4.92
45-64 years	97.00	3.00
65 + years	97.75	2.25

Table 2. Comparison of selected characteristics of adults (18 years and older) living in telephone and nontelephone households in the U.S., 2002 NHIS

Selected Characteristics	Total (%)	Test Sample from Telephone Households (%)	Telephone Households (A) (%)	Nontelephone Households (B) (%)	Difference (B - A) (%)
Age 18-44	46.55	45.78	45.55	64.97	19.42
Age 45-64	33.69	34.68	34.07	26.67	-7.4
Age 65+	19.75	19.54	20.37	8.35	-12.02
Non-Hispanic Whites and Others	78.60	79.02	79.36	64.62	-14.74
Hispanics	9.52	9.24	9.23	14.85	5.62
Non-Hispanic Blacks	11.89	11.73	11.42	20.53	9.11
Male	53.79	53.48	53.63	56.67	3.04
Female	46.21	46.52	46.37	43.33	-3.04
Owner	66.18	68.39	68.30	27.25	-41.05
Renter	31.73	30.09	29.94	64.56	34.62
Income > $20,000	16.16	17.42	16.62	7.68	-8.94
Income <= $20,000	5.71	5.34	5.42	11.11	5.69
Health Status: Excellent, Very Good	60.49	61.59	61.15	48.46	-12.69
Health Status: Good	25.77	24.89	25.47	31.35	5.88
Health Status: Fair, Poor	13.14	13.20	12.90	17.46	4.56
Has Private Health Insurance	69.91	71.54	71.73	36.55	-35.18
No Private Health Insurance	29.16	27.45	27.45	60.47	33.02
No Health Insurance at all	14.26	13.17	13.01	37.14	24.13
Has Health Insurance	84.81	85.82	86.17	59.88	-26.29
Has Limitation of Activity	16.95	16.13	16.80	19.63	2.83
No Limitation of Activity	82.69	83.52	82.87	79.44	-3.43
Education: Less Than High School	43.40	42.36	42.64	57.34	14.70
Education: Some College	44.45	45.28	45.30	28.93	-16.37
Education: Graduate School	8.97	9.63	9.31	2.70	-6.61

Table 3. Comparison of percentage of adults (18-64 years) living with at least one elderly in telephone or nontelephone households in the U.S., 2002 NHIS

Selected Characteristics	2002 NHIS			
	All Households (%LWE)	Telephone Households (% LWE)	Nontelephone Households (% LWE)	Difference in %LWE (Nontelephone - Telephone)
All adults18-64 years	4.52	4.66	1.76	-2.90
No Telephone	1.76	-	1.76	-
Has Telephone	4.66	4.66	-	-
Age 18-44	2.66	2.74	1.45	-1.29
Age 45-64	6.85	7.00	2.50	**-4.50**
Non-Hispanic White	3.86	3.97	1.26	-2.71
Hispanic	6.45	6.72	2.89	**-3.83**
Non-Hispanic Black	5.35	5.66	1.86	**-3.80**
Non-Hispanic Others	6.84	7.08	2.77	**-4.31**
Male	3.37	3.48	1.22	-2.26
Female	6.07	6.26	2.45	**-3.81**
Owner	5.46	5.50	3.60	-1.90
Renter	2.42	2.60	0.81	-1.79
Income > $20,000	5.99	6.10	1.78	**-4.32**
Income <= $20,000	6.35	6.94	2.18	**-4.76**

Table 4. Comparison of selected estimates using the simple poststratification to adjust for noncoverage of nontelephone households with those using model-based weighting class Methods 1 or 2 in the test sample

Characteristics	Difference (Method 1 - SP) (%)	Std. Error Ratio (Method 1 / SP) (%)	Difference (Method 2 - SP) (%)	Std. Error Ratio (Method 2 / SP) (%)
Health Status: Excellent, Very Good	-0.34	1.01	-0.28	1.00
Health Status: Fair, Poor	0.13	1.02	0.15	1.01
No Private Health Insurance	1.01	1.04	1.02	1.03
No Health Insurance	0.75	1.04	0.74	1.04
Has Limitation of Activity	-0.01	1.01	0.01	1.00
Income Less Than $20,000	0.37	1.05	0.27	1.05
College Graduates	-0.43	0.96	-0.43	0.97

Table 5. Comparison of the 2002 NHIS estimates with the estimates using a simple poststratification method to adjust for noncoverage of nontelephone households in the test sample

Characteristics	2002 NHIS		Test Sample (SP Method)		Difference SP - NHIS	Std. Error Ratio (SP / NHIS)
	Estimate (%)	Std Error (%)	Estimate (%)	Std Error (%)	(%)	(%)
Health Status: Excellent, Very Good	60.49	0.33	63.23	0.72	2.74	2.18
Health Status: Fair, Poor	13.14	0.22	12.24	0.46	-0.90	2.09
No Private Insurance	29.16	0.34	27.49	0.73	-1.67	2.15
No Health Insurance	14.26	0.23	13.93	0.59	-0.33	2.56
Has Limitation of Activity	16.95	0.28	14.92	0.54	-2.03	1.93
Income Less Than $20,000	5.71	0.17	5.32	0.36	-0.39	2.12
College Graduates	8.97	0.21	9.46	0.50	0.49	2.38

Table 6. Percent difference in the final weighted estimates computed from the JCUSH/U.S.A. data and the NHIS 2002 data by age group

Characteristics	Age Group		
	18-44years Difference JCUSH - NHIS (%)	45-64 years Difference JCUSH – NHIS (%)	65+ years Difference JCUSH – NHIS (%)
Health Status: Excellent, Very Good	**-7.4**	0.2	2.6
Health Status: Fair, Poor	3.0	1.3	2.1
No Private Insurance	**-5.0**	-4.4	-0.9
No Health Insurance	**-5.9**	-3.5	0.1
Income Less Than $20,000	**-5.6**	-2.0	-4.0
College Graduates	3.6	2.2	5.2