SOCIAL CORRELATES OF RESPONSE BIAS IN THE NATIONAL HEALTH INTERVIEW SURVEY: LESSONS FOR THE REDESIGN

Felicia B. LeClere and P. Ellen Parsons, National Center for Health Statistics
Felicia B. LeClere, National Center for Health Statistics, 6525 Belcrest Road -Room 850
Hyattsville, MD 20782

Key Words: Survey, Response Bias, NHIS

The basic core questionnaire for the National Health Interview Survey (NHIS) is being redesigned for administration in 1996. One of the ongoing criticisms of the NHIS and other interview instruments used to gather data about health and well-being is that there is differential non-response and response bias by the socioeconomic characteristics of the respondents especially race, ethnicity and income (Salber and Beza, 1980; Vernon, Roberts, and Lee, 1984). Overall non-response, item non-response, and inaccuracy in detailed recall are said to complicate comparisons made between demographic groups and the inferences made about subpopulation differences in health and well-being between groups (Johnson et al., 1994). One of the tasks of the NHIS redesign committee is to address these on-going criticisms by re-examining both the variability in response patterns for population subgroups and the quality of the socio-demographic information itself.

In this paper, we address two aspects of this more general analytical problem by examining the socio-demographic correlates of overall and item non-response from the 1990 Family Resources supplement to the NHIS(NHIS-FR) and measures of agreement between selected health items from the 1991 NHIS reinterview program. In the first portion of this paper, which is based on the NHIS-FR, existing measures of social status and socioeconomic characteristics are examined to identify possible bias in reporting patterns by sub-population characteristics. Our major concern is whether the questions related to economic characteristics are themselves significantly biased by the pattern of non-response and, in particular, among those of lower socioeconomic status and for minority group members. The second portion of the paper, using the reinterview data, more directly addresses the problem of socio-demographic variability in response to health items. In the final section, we draw some implications from this analysis for the redesigned questionnaire.

Measuring socioeconomic status

The socio-demographic questions on the National Health Interview Survey have remained relatively stable since its inception in 1957. Most of the modification has occurred in questions about race and ethnicity rather than in direct measures of socioeconomic status. The core NHIS questionnaire now contains questions on race, ethnicity, individual educational attainment and that of the responsible adult in the household, categorical income questions which are used to construct poverty status, labor force status, and characteristics of current job for adult members of the household. Other demographic information includes age, sex, household composition, marital status, living arrangements, duration of residence, place of origin for immigrants, veteran's status, and some characteristics of the place of residence. Simple static measures of social status and current measures of economic well-being can be generated from these items for individuals and families in the sample. No information about the sources of income, health insurance status, or receipt of transfer income is available.

Beginning in 1990, a permanent supplement was added to the questionnaire to redress what were perceived to be severe short-comings in the economic information collected on individuals and households in the NHIS. The Family Resources supplement contains questions about employment characteristics, sources of family income, Federal program participation and health insurance coverage. The underlying premise for adding this permanent supplement was that the complex relationship between income, employment, and the health care system in the United States implies that discretionary income and sources of income as well as employment have important implications for differential health care use and consequent health.

For the first time, based on the experiences of other Federal data collection activities, the NHIS began to collect data on discrete dollar amounts for particular sources of income. Experience with detailed categorical income questions suggested that item non-response, which is approximately 20 percent of the respondents annually for that question, would be quite high for many of the continuous income items. Item and questionnaire imputation was the suggested remedy for what was anticipated to be high and differential non-response. The method of imputation put in place for the 1990 NHIS-FR, which is the only year of data from the family resources questionnaire currently available, is similar to the hot deck method currently used by the

Current Population Survey conducted by the Census Bureau. The data file contains flags that can be used to identify the individual data records that have been imputed. We use these flags to conduct analyses of non-response for the income and program participation items.

The analysis in this section is in two parts. First, we examine, in cross-tabular format, imputation rates for the overall questionnaire and 11 of the yes/no items by a variety of socioeconomic and demographic characteristics. This analysis is based on the entire sample using sample weights to adjust for non-response and differential oversampling. In the second piece of this analysis, we use multivariate models to examine the determinants of imputation for two money items, monthly salary receipt and interest income, which have high item imputation rates as well as all item imputation. In an ancillary analysis, we also examine the determinants of imputation of the amount of monthly welfare receipt, although it is relatively low, to see if low income populations have different response patterns. These two analyses are based on adult self-responders only.

Using imputations rates to assess the nature and impact of differential non-response bias is problematic in several respects. First, records were imputed if the respondents answered don't know or refused to answer the question or the data editing procedure turned up flawed, inconsistent or inappropriate codes. Many researchers have shown that different types of cognitive processes, and thus, underlying population response patterns, generate these types of answers (Jabine et al., 1984). Thus, an imputation rate is not a clean measure of the concept of non-response but rather represents an amalgam of response problems. Second, these analyses exclude those who refused to participate in the survey or could not be located for interview. The survey weight includes an adjustment for non-response but is calibrated for only a limited number of characteristics. Therefore, this analysis excludes the most damaging forms of non-response and consequent bias.

Imputation Rates

Results from the descriptive analysis of imputation rates are presented in Table 1 (available from the authors). The first column, contains imputation rates for those records on which all items on the NHIS-FR have been imputed. This represents the population who simply refused to continue the interview when the interviewer reached this portion of the questionnaire. The remaining columns contain imputation rates for items that require yes/no answers about receipt of particular types of income and participation in Federal programs in the past month.

The all-item imputation rates follow a demographic pattern that could be expected for refusals. Blacks, persons 25 to 64 years of age, the never married, those living alone, and residents of the West and in central cities as well as those who refused to answer questions on the basic NHIS are more likely to have all items on the NHIS-FR imputed. This is consistent with past representations of the population of non-responders who refuse to answer all or part of the questionnaire. Standard measures of socio-economic status, on the other hand, fail to differentiate respondents and non-respondents. Respondents with less than a high school education have similar imputation rates to those with a college education. Income also is not related to non-response in a predictable manner. Slightly elevated rates occur among the wealthy and the non-poor. For all groups, the all-item imputation rates, which run between 2 and 4 percent, are consistent with overall survey non-response rates for the NHIS (Benson and Marano, 1994).

The demographic pattern of item imputation rates offer interesting insight into both the nature of the questions and the saliency of those questions to groups of respondents. Persons who refuse or do not know whether they receive Medicare are more likely to be old, female, poor, previously married, and poorly educated. Because this question addresses receipt of Medicare benefits in the last month, it is likely the case that the majority of these persons simply did not know about their eligibility in the previous month. There are very low imputation rates for members of the population who are very unlikely to be eligible for Medicare. Welfare and Food Stamp receipt show little of this variation, with among the lowest imputation rates of all the items shown. Even those in the population who are more likely to be eligible, including young, low income, female, and minority populations the imputation rates by item remain below 1 percent consistently. This comparisons suggest that Medicare eligibility and participation may be less clear among the eligible population than other Federal programs.

The source of income items are, perhaps, even more revealing. For items such as pension allotments, disability pensions, and social security payments, which can be expected to be disbursed monthly or bimonthly, the imputation rates are low overall and not consistently related to demographic characteristics. Conversely, interest and dividend income are more likely to be imputed and are consistently related to socio-demographic characteristics. There is a unexpectedly sharp gradient by age, income, and education with middle-aged, highly educated and wealthier persons more likely to have imputed values for receipt of

interest and dividends in the previous month. Imputation rates are as high as 5 1/2 percent for some subgroups. This is counterintuitive for simple 'yes/no' questions. Although we cannot know definitively without separate don't know and refused responses, it appears from the demographic patterns, that the respondents are likely to receive such income but are unsure whether they received it last month since interest and dividend income is not usually disbursed in monthly allotments.

Sociodemographic Correlates of Imputation

The association between demographic correlates and non-response appears to depend on the saliency of the question to the subpopulation and whether the non-response can be construed as a refusal or don't know. To sort out the conditional association between respondents' characteristics and imputation, we construct multivariate logistic regression models to predict the likelihood that an item will be imputed given the characteristics of the respondents. In table 2a, we examine three items including whether all items were imputed for the respondent and two items with high overall imputation rates, monthly earnings for those employed by others and amount of monthly interest income. The independent variables are dummy variables, which mirror those used in the Table 1. The categories not shown are the omitted categories. We added several health items to this analysis in order to assess the confounding influence of non-response in analyses of the relationship between health and socioeconomic status as measured from the NHIS-FR. Persons in poor health and those with activity limitations are compared to their healthly counterparts without health-related limitations.

The model for all item imputation is for the most part consistent with the descriptive results in Table 1. Race, living arrangements, sex, age, and education and those living in non-metropolitan areas and in the West are significantly associated with having imputed records. Blacks and those living in central cities are 40 to 70 percent more likely to have their answers imputed relative to their non-black and non-metro counterparts. There is also an association with health and non-response. Those in poor health are almost 20 percent more likely to have the all items imputed on the NHIS. There is no association between poverty and Hispanic ethnicity and imputation.

The salary and interest models show some consistency with the descriptive models and reinforce the idea that salience might also be critical to imputation for dollar amounts as well as 'yes/no' items. The interest model shows significantly lower probability of imputation among males, the poorly educated, those living in central cities, minority populations, those in poor health and those with activity limitations. These models are estimated only for persons who receive interest income, which implies that the imputation relates to an inability or unwillingness to name a dollar amount. Without further information, the results of this model suggests by inference from the demographic characteristics of those who are more likely to have complete records that where interest income is small, the exact dollar amount may be quite easy to remember. On the other hand, those of high income, which is measured by family income greater the 50,000 dollars per annum, for whom interest income may represent only a small portion or a variable contribution to total income, the odds that their record will be imputed is nearly 80 percent greater than those with lower incomes.

Table 2b presents further evidence that the pattern of non-response, as measured by the probability of imputation, is a reflection of both the saliency and strictures relevant to the item as they relate to the population subgroup. The dependent variable in this model is the probability of item imputation for the dollar amount of monthly welfare income for those who acknowledge receiving welfare and whose answer was not imputed for the yes/no receipt of welfare. The significant variables reflect welfare eligibility. Males, married persons, and people in the labor force are between 60 and 80 percent more likely to have an imputed amount in monthly welfare receipt field than their female, unmarried counterparts who are out of the labor force. The poor conversely are half as likely to have an imputed amount. From our knowledge of eligibility for welfare benefits, the non-responders in this case are likely to be people who refuse to give dollar amounts because they fear reprisal.

Measuring health

The most important criticism leveled at interview instruments of health is that there is considerable bias in how different subgroups of the population respond to particular types of questions about health (Angel and Guarnaccia, 1989). The variability in response patterns is then misinterpreted as substantive differences in health rather than measurement error. In this section, we examine relatively straightforward measures of agreement for basic questions about health care use and health behavior.

In the final part of this analysis, we use a different data source to assess a respondent's ability to reproduce answers when asked the same question again. For the 1991 data year, the NHIS field staff reinterviewed 4,376 persons in 1,711 households by telephone for selected questions from the 1991

supplements. We use three questions from questionnaires that were administered to a randomly selected adult per household. The three questions from the reinterview are as follows: (1) Have you smoked at least 100 cigarettes in your lifetime? (2) Has a doctor EVER told you that you have hypertension or high blood pressure? (3) Do you have a place that you go to when you are sick or need health care on a regular basis?

These questions were asked of all participants in the reinterview, thus, we retain the limited amount of sample available. In addition, these questions are much less likely to have an extremely skewed distributions which causes measures of agreement to be artificially inflated (Edwards, Winn, Kurlantzick, et al., 1994).

In Table 3 (available from the authors), these three questions are cross-tabulated with two measures of agreement between time 1 and time 2 responses. The column 'percent common' is simply the proportion of persons in that demographic group that gave the same response in time 1 as they did in time 2. The proportion with common agreement is not a particularly good measure of reliability because respondents could have easily given the same response based on chance. The Kappa statistic, used for studies that compare survey responses to data from records, allows for the possibility that the agreement is by chance alone. Kappa is also flawed for reinterview studies because it assumes that at least one of the responses is measured without error, which is clearly not the case in the study (Biemer, 1993). Nevertheless, it remains adequate for descriptive purposes. We drop from the calculation of Kappa, cases that were unknown in either time 1 or time 2 because of their artificial effects on the statistic.

Like our analysis of non-response, this analysis shows more variability by question type than it does by demographic group. The smoking and hypertension questions show relatively high levels of agreement and Kappa statistics in the upper ranges across almost all demographic groups. Blacks and those living in the Northeast are likely to have lower levels of agreement on these two questions but still in the range of moderate agreement. The effect of education is as expected for the smoking question with more educated people more likely to respond consistently.

The question about regular source of care is more problematic. Although proportions with common response are high, the Kappas are in the low range across all demographic groups. Because there was a time lag between the first and second interviews, it may be that for some participants, their interpretation of the regular source of care changed. The subjective nature of this question, however, suggests that many respondents change their interpretation of the concept of

'regular source of care' instead of their use of such care.

In summary, then, the original premise that many of the substantive results of health surveys for demographic subgroups are a consequence of measurement error is unfounded. The demographic consistency within questions and the variability across questions suggest that measurement error affects subgroups in the same fashion.

Lessons for the Redesign
This research has lead to some of the following conclusions for the redesign: (1) Questions about sources of income need not be consistent in reference period with each other but rather with the nature of the income source. Interest and dividend income had relatively high levels of imputation largely because of the reference period problem. (2) Item non-response in income questions is related more to the saliency of the question to the demographic subgroup than to a standard demographic model of non-response. (3) For eliciting truthful answers about dollar values from those that receive Federal transfer income, it is important to improve the perception of anonymity so that those who may not be eligible feel safe to give truthful answers. (4) Measurement error in health questions is more consistently related to the nature of the question than to subpopulation variability in response patterns.

Full paper and tables available from authors.

Angel R. and Guarnacci PJ. 1989. Mind, body, and culture: somatization among Hispanics. Social Science and Medicine. 28(12):1229-1238.

Benson V and Marano M. 1994. Current Estimates from the National Health Interview Survey, 1992. National Center for Health Statistics: Vital and Health Stat 10(189).

Biemer P. 1993. Reinterview:discussion. InProceedings of the Bureau of the Census 1993 Annual Research Conference. Washington, D. C.: Government Printing Office, pp. 739-742.

Edwards WS, Winn DM, Kurlantzick V et al., 1994. Evaluation of National Health Interview Survey Diagnostic Reporting. National Center for Health Statistics. Vital Health Stat 2(120).

Jabine TB, Straf ML, Tanur JM and Tourangeau R. 1984. Cognitive Aspects of Survey Methodlogy: Building a Bridge Between Disciplines. Washington, DC: National Academy Press.

Johnson J, O'Rourke D, Chavez N, Sudman S, Warnecke R, Lacey L, Horm J. 1994. Development of a methodology for examining cultural differences in the cognitive processing of health survey questions. In Proceedings of the Conference on Research Issues to Address Minority Health Data Needs. Hyattsville, MD: National Center for Health Statistics.

Salber EJ and Beza AG 1980. The health interview survey and minority health. Meidcal Care. 28:319-326.

Vernon SW, Roberts RE, and Lee ES. 1984. Ethnic status and participation in longitudinal health surveys. American Journal of Epidemiology. 119:99-113.

Table 2a: Determinants of imputation rates for self-reporting adults: logistic regression

	Item					
	All items		**Salary**		**Interest**	
Determinant	Coeff[1]	Odds	Coeff	Odds	Coeff	Odds
Intercept	**-3.82**		**-6.35**		**-2.71**	
Age	**-.01**	.99	**.01**	1.01	**.03**	1.03
Male	**.09**	1.09	**-.12**	.89	**-.20**	.82
Married	**.20**	1.22	**.19**	1.21	**.35**	1.42
< high school	**-.72**	.49	**-.15**	.89	**-.53**	.59
High school	**-.59**	.55	.02	1.02	**-.12**	.89
College	**-.55**	.58	.03	1.03	**.10**	1.11
Living alone	**.35**	1.42	-.05	.95	**.34**	1.41
Employed	**.21**	1.23	**3.68**	39.8	.07	1.07
North	-.08	.92	**.40**	1.49	**.26**	1.29
Midwest	.00	1.00	**.25**	1.28	**.16**	1.17
West	**-.76**	.47	**.20**	1.22	**.12**	1.13
Central city	**.53**	1.69	**-.19**	1.22	**-.11**	.90
Suburbs	**.50**	1.64	-.08	.92	.02	1.02
Black	**.39**	1.47	**.35**	1.41	**-.60**	.55
Poor	.12	1.13	.04	1.04	**-.83**	.44
Unknown income	**1.87**	6.46	**1.85**	6.38	**.35**	1.43
Upper income	-.13	.88	.06	1.06	**.57**	1.76
Hispanic	.12	1.13	**-.29**	.75	**-.45**	.64
Poor health	**.17**	1.19	-.04	.96	**-.21**	.81
Disabled	-.12	.89	.01	1.01	**-.20**	.82

[1] Coefficients in bold are significant at the $p \le .05$ level

Table 2b: Determinants of imputation rates for amount of welfare receipt for those receiving welfare, self-reporting adults: logistic regression

Determinant	Monthly Welfare Receipt	
	Coefficient[1]	Odds
Intercept	**-3.02**	
Age	**.02**	1.02
Male	**.59**	1.81
Married	**.55**	1.73
< high school	.52	1.05
High school	.29	1.33
College	.52	1.77
Living alone	.28	1.32
Employed	**.46**	1.59
North	-.15	.87
Midwest	.15	1.12
West	-.00	.99
Central city	-.21	.81
Suburbs	**-.77**	.46
Black	.14	1.15
Poor	**-.63**	.53
Unknown income	**1.07**	2.91
Upper income	.73	2.01
Hispanic	.27	1.32
Poor health	.06	1.06
Disabled	.36	1.43

[1] Coefficients in bold are significant at $p \leq .05$ level