Assessing Oversampling of Older Persons in a National Health Survey

Karen E. Davis (kedavis@cdc.gov) and Chris Moriarity (cmoriarity@cdc.gov)
National Center for Health Statistics, Centers for Disease Control and Prevention,
3311 Toledo Road, Hyattsville, MD 20782

Key Words: Sample survey, oversampling

1. Introduction[1]

The National Health Interview Survey (NHIS) is one of the major data collection programs of the National Center for Health Statistics (NCHS). The sample design for the NHIS traditionally has undergone a redesign about every 10 years to address new and continuing data needs at the national and subnational levels, and for subpopulations of interest. The ability to produce reliable annual estimates for older persons age 65 years and over by race and ethnicity was a major design objective in the most recent NHIS sample redesign research. In 2002, research was conducted for the 2006-2014 NHIS redesign to assess options for oversampling older minority persons (Davis, 2003). Beginning with the 2006 NHIS, when a new sample design was implemented, minority persons age 65 years and older have an increased probability of selection as sample adults. The sample adult is one family member who is randomly selected to provide basic information on health status, health care services, and behavior. This individual responds for him/herself (no proxy) unless he/she is physically or mentally unable to do so. This paper describes the research that has been conducted using the 2006 NHIS to assess the selection and sample yield for older minority persons from the revised sampling procedure. Minority persons are defined as Hispanics, non-Hispanic Asians, and non-Hispanic blacks in this paper.

2. Background

The National Health Interview Survey (NHIS) is a multi-purpose health survey conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), and is the principal source of information on the health of the civilian, noninstitutionalized, household population of the United States. The NHIS has been conducted continuously since its beginning in 1957. The data collected in the NHIS are obtained through a complex sample design involving stratification, clustering, and multistage sampling. During the 1995-2005 sample design period, both the non-Hispanic (NH) black and Hispanic populations were oversampled to allow for more precise estimation of health in these growing minority populations. Areas with higher concentrations of NH blacks and Hispanics were sampled at a higher rate. All households containing any eligible NH black or Hispanic persons were selected for interview, while only a subsample of the other households were selected for interview.

NCHS initiated a redesign of the NHIS that was implemented in 2006. The new sample design included several expansions in oversampling procedures. Beginning with the 2006 NHIS, all households containing any eligible NH Asian person also are selected for interview. The sample adult (SA) selection process has also been revised so that when Hispanic, NH black, or NH Asian persons age 65 years or older are present in the household, they have an increased chance of selection as the sample adult (NCHS, 2007).

The previous design could produce annual estimates for NH blacks and Hispanics with satisfactory precision (i.e., estimate has relative standard error less than 30%). However, equivalent precision was not attained for subdomains

---

1 The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

defined by sex and age within a race-ethnic group (e.g., Hispanic males and females 75+ years of age) (Davis, et al., 2001). The ability to produce reliable annual estimates for the elderly population by race and ethnicity, within the age groups 65-74 and 75+, was a major redesign objective (Ezzati-Rice, et al., 2001).

In 2002, research was conducted to explore the effect of increasing the probability of selection of sample adults age 65 years and older in minority households (Davis, 2003). As expected, the results indicated that the nominal sample sizes for sample adults age 65 years and older increased as the selection probabilities increased, with a corresponding decrease in design effect relative to a simple random sample of the population. At NCHS' request, the Census Bureau staff completed several tests to simulate the selection process in oversampling minority sample adults age 65 years and older. The Census Bureau's Random Sample Adult Generator selection process provides a weight of 1.0 to each sample adult in a family, except for a minority person 65+ years old, who receives a weight of 2.0 (doubling the chances of being selected). In testing the accuracy of the random number generator, the actual results came very close to theoretical figures, indicating that the Random Sample Adult Generator is performing as intended with minor fluctuations attributed to sampling variability (Davis, et al., 2006). Sample adults were selected in the simulations in the proper proportions, taking account of sampling variability during the simulations.

3. Methods

For this research, we used data from the 1997 and 2006 National Health Interview Surveys. The NHIS Person-Level file contains information for all household members. With the 1997 NHIS Person-Level file, we focused on households containing minority persons, categorizing the household members by age 65 years and older (elderly) and younger than age 65 years (non-elderly). We assumed each household was one family to simplify the estimation process. Since the 1997 interviewed sample contained 39,832 households with 40,623 families, we thought that assuming one family per household was a reasonable approximation. Initially, we assigned each adult the same probability of selection as the sample adult. Next, the elderly were given a higher chance of selection by doubling their sample count, and we assigned revised selection probabilities based on these counts. Then the expected yield of elderly and non-elderly persons was calculated using SAS software to obtain the sample sizes for this "doubling" sampling protocol. Table 1 provides the expected yield for the total sample, the expected yield of elderly and non-elderly persons, and the estimated design effects. Note that the original estimates in (Davis, 2003) assumed each family member was the same race/ethnicity. Improvements were made to the original estimates to allow for families containing members of different race/ethnic groups (Hispanic, NH Asian, NH black, and NH other). In 1997 there were about 1900 such households.

The design effects ($DEFF$) were calculated to provide an indication of the relative increase in variance due to the unequal sample adult sampling rates (e.g., Potter, 1988). $DEFF = n \sum w_i^2 / (\sum w_i)^2$, where w_i is the sampling weight for the ith unit and n is the sample size. The results in Table 1 indicate that the expected nominal sample sizes for the elderly increase as the selection probabilities increase, with a corresponding decrease in design effect for the elderly. Note that the expected yields do not include an adjustment for nonresponse. In 1997, the conditional response rate for sample adults was 89%. That is, given that a family responded to the NHIS, about 89% of the time a sample adult also responded.

We used the 1997 and 2006 NHIS Sample Adult files to calculate the total number of elderly and non-elderly persons by race and ethnicity. For 2006 estimates, we used all sample adult persons in minority groups, regardless of the household composition, with variables "origin_i" and "racreci3" to assign race/ethnicity group. (These variables are not available in the 1997 NHIS data files. They were used to assign race/ethnicity in the 2006 NHIS weight creation process.) Note that the 2006 conditional response rate for sample adults was 81.4%. The sample adult response rates have declined over time due to a number of issues such as increasing refusals and noncontacts.

4. Results

In Table 1, note that the expected yields do not reflect nonresponse to the sample adult module, but show the available nominal sample. Table 2 lists the 2006 NHIS SA counts, with 2006 being the first year of the new NHIS sample design. Note that there are large differences between the overall 1997 and 2006 sample sizes, yet the sample percents by subgroup are very similar. In 2006, the survey was faced with a budget shortfall. As a result, the size of the 2006 NHIS sample was reduced by approximately 50% during July-September 2006, a 1/8 reduction overall for the year. This cutback was in addition to the ongoing overall 12.5% reduction due to the new sample design in 2006 (NCHS, 2007). In Table 2, we also calculated the design effects for each race/ethnicity subgroup. Note that the design effects in Table 2 are more realistic than those in Table 1. The design effects in Table 2 are from actual sample adult outcomes, whereas the design effects in Table 1 are based on estimated yields, which included the selection of fractions of persons in households with multiple race/ethnic groups, something that of course is unattainable in actual practice, leading to artificially low design effects. Table 3 lists the actual 1997 NHIS SA counts and the corresponding design effects. Overall, there was a 7.4% decrease in the number of minority sample adults age 65 years and over, yet the NH Asian sample of persons age 65 years and over increased by more than 60% in the new design. Due to the revised sample adult selection protocol, the variability for the elderly minority sample adult weight factors is lower for 2006 than 1997, and vice versa for the non-elderly minority factors. As expected, we see a decrease in the elderly minority design effects in Table 2 when compared to the design effects in Table 3, and vice versa for the non-elderly minority design effects. In Table 4 we inflated the 2006 NHIS SA counts (by multiplying by 1/.875) to adjust for the sample reduction due to the 2006 budget shortfall, in order to see what the potential minority sample sizes would have been. Note that we did not adjust for the overall reduction due to the new sample design since that reduction is immutable. We compared these counts with the actual 1997 NHIS SA counts in Table 3. We found that if the 2006 budget shortfall had not taken place, we would have seen a 6% increase for all sample adults age 65 years and over. This increase would have been solely due to the 66% increase in the number of NH Asians that were oversampled for the first time in the 2006 NHIS. Table 4 seems to indicate that if the 2006 budget shortfall had not occurred, the sample counts in the first year of the new design would still have fewer elderly Hispanic and NH black sample adults than in 1997.

5. Discussion

Despite sample reductions in the 2006 NHIS file, the NH Asian sample adult counts have increased by about 32% overall from 1997. Elderly NH Asian counts have also increased as compared to 1997 counts. These increases in sample yield for NH Asians were predicted in other 2006-2014 redesign research (Jones, et al., 2002). The combination of retaining (or "screening in") NH Asian persons in households within each sample segment and doubling the chance of selection of minority persons age 65 years and over both have a demonstrated effect. We plan to continue to monitor future NHIS data to see how the new design impacts sample adult estimates, although unfortunately the 2007 NHIS has also undergone a sample reduction due to a budget shortfall. The intent of the new design was to increase the sample size for all elderly minorities, but this has not been achieved yet. The increase in the NH Asian sample size in the new sample design, was realized even with a reduced overall sample size for the NHIS.

References

Davis, K. E., Ezzati-Rice, T., Gonzalez, J.F., Jones, C., Moriarity, C., Tompkins, L., *2005-2014 NHIS Sample Redesign Research Report: Current Precision of Selected Health Statistics*, NCHS Internal Report, May 15, 2001.

Davis, K., (2003). *Two Options for Oversampling in the National Health Interview Survey.* Proceedings of the 2003 Annual Meeting of the American Statistical Association [CD-ROM], Alexandria, VA: American Statistical Association.

Davis, K., Moriarity, C., (2006). *Assessing Population Coverage in a Health Survey.* Proceedings of the 2006 Joint Statistical Meetings [CD-ROM], Alexandria, VA: American Statistical Association.

Ezzati-Rice, T., Moriarity, C., Katzoff, M., Parsons, V., (2001). *Overview of Sample Design Research for the National Health Interview Survey,* Proceedings of the 2001 Annual Meeting of the American Statistical Association [CD-ROM], Alexandria, VA: American Statistical Association.

Jones, C., Moriarity, C., Davis, K., Gonzalez Jr., J.,(2002). *Methods to Improve the Precision of Health Statistics for Non-Hispanic Asians in the National Health Interview Survey,* Proceedings of the 2002 Annual Meeting of the American Statistical Association [CD-ROM], Alexandria, VA: American Statistical Association.

National Center for Health Statistics (2000). *Data File Documentation, National Health Interview Survey, 1997 (machine readable data file and documentation).* National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland.
ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/1997/srvydesc.pdf

National Center for Health Statistics (2007). *Data File Documentation, National Health Interview Survey, 2006 (machine readable data file and documentation).* National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland.
ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2006/srvydesc.pdf

Potter, F.,(1988) *Survey of Procedures to Control Extreme Sampling Weights,* Proceedings of the 1988 Annual Meeting of the American Statistical Association, Alexandria, VA: American Statistical Association.

Table 1. Expected Number and Percent of Sample Adults for Households Containing Minorities with Double Probability of Selection for Elderly Minorities, 1997 NHIS

Race/Ethnicity	Total Sample*	Elderly Yield	Pct	Non-Elderly Yield	Pct	Total Design Effect	Elderly Design Effect	Non-Elderly Design Effect
All 3 Minorities	13,164	1,709		11,455				
Hispanic	6,411	672	10%	5,739	90%	1.059	0.890	1.084
NH-Asian	1,019	114	11%	905	89%	1.015	0.853	1.039
NH-Black	5,734	923	16%	4,811	84%	1.118	0.929	1.161

*Note: The total sample column shows the expected nominal sample and does not reflect nonresponse to the sample adult module.

Table 2. Number of Sample Adults for Households Containing Minorities, 2006 NHIS

Race/Ethnicity	Total Sample	Elderly Sample	Pct	Non-Elderly Sample	Pct	Total Design Effect	Elderly Design Effect	Non-Elderly Design Effect
All 3 Minorities	9,551	1,302		8,249				
Hispanic	4,227	422	10%	3,805	90%	1.241	1.131	1.243
NH-Asian	1,306	206	16%	1,100	84%	1.275	1.167	1.286
NH-Black	4,018	674	17%	3,344	83%	1.253	1.138	1.267

Table 3. Number of Sample Adults for Households Containing Minorities, 1997 NHIS

Race/Ethnicity	Total Sample	Elderly Sample	Pct	Non-Elderly Sample	Pct	Total Design Effect	Elderly Design Effect	Non-Elderly Design Effect
All 3 Minorities	11,664	1,398		10,266				
Hispanic	5,685	530	9%	5,155	91%	1.227	1.258	1.222
NH-Asian	892	80	9%	812	91%	1.267	1.256	1.268
NH-Black	5,087	788	15%	4,299	85%	1.241	1.230	1.241

Table 4. Projected Sample Adult Yields for Households Containing Minorities, Adjusted to Account for Sample Reduction, 2006 NHIS

Race/Ethnicity	Total Sample	Elderly	Non-Elderly	Total %Chg*	Elderly %Chg*	Non-Elderly %Chg*
All 3 Minorities	10,915	1,488	9,427	-6.9	6.0	-8.9
Hispanic	4,831	482	4,349	-17.7	-9.9	-18.5
NH-Asian	1,493	235	1,257	40.2	66.0	35.4
NH-Black	4,592	770	3,822	-10.8	-2.3	-12.5

* Note: % change between 2006 NHIS sample adult adjusted estimates and 1997 NHIS sample adult estimates (Table 3).