

## Assessing Population Coverage in a Health Survey

Karen E. Davis ([kedavis@cdc.gov](mailto:kedavis@cdc.gov)) and Chris Moriarity ([cmoriarity@cdc.gov](mailto:cmoriarity@cdc.gov))  
 National Center for Health Statistics, Centers for Disease Control and Prevention, 3311 Toledo Rd,  
 Room 3213, Hyattsville, MD 20782

### Abstract

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 approximately every 10 years to address new and continuing data needs at both the subnational level and for minority and economic subdomains of the population. The ability to produce reliable annual estimates for the population by race and ethnicity, within the age groups 65-74 and 75+, is a major design objective. In 2002, research was conducted to assess several options for oversampling sample adult (SA) minority persons age 65 years and older to meet the goal of improving the precision of estimates for them, while retaining close to the same precision for other estimates and keeping the overall sample size constant (Davis, 2003). Beginning with the 2006 NHIS, the option to retain the current protocol of selecting one sample adult per household, but giving minority persons age 65 years and older an increased probability of selection, is being implemented. This paper describes the research that has been conducted to assess the population coverage for these sample adults, to analyze the effects of the increased oversampling, and to give a picture of the expected increase in sample yield for minority persons age 65 years and older.

**Keywords:** sample survey, oversampling, population coverage

### 1. Introduction

This paper describes research in oversampling minority (Hispanic, black, and Asian) persons age 65 years and older. The overall goal is to improve estimates for these selected subgroups, without significant adverse effects to other important estimates. With this methodology, we assumed a fixed budget, so sample size remains fixed.

### 2. Background

In 1997, many changes were made to the NHIS interview protocol. The biggest change was that computer-assisted personal interviewing (CAPI) replaced paper questionnaires. Another major change is that a relatively short list of questions are asked

about most household residents. One sample adult is randomly selected from each family, and that sample adult is asked a number of questions that previously were asked about all adults in the family. If any children are in the family, one sample child is randomly selected, and questions are asked about that child that previously were asked about all children in the family.

Prior to 2006, the procedure for selecting the sample adult from a family did not involve any random reshuffling of the roster prior to sample selection. A random reshuffle was not necessary, given that all sample adults had the same chance of being selected. Our research included an examination of whether assigning differential selection probabilities for some adults also required a random reshuffling of the roster prior to sample selection. Our findings are described below.

### 3. Data

For this research, we used data from seven years (1997-2003) of the National Health Interview Survey (NHIS) and the Current Population Survey (CPS). The 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. It is an annual survey, and data are collected through in-person interviews (NCHS, 2004).

The data collected in the NHIS are obtained through a complex sample design involving stratification, clustering, and multistage sampling. Both the black and Hispanic populations are oversampled to allow for more precise estimation of health in these growing minority populations. Areas with higher concentrations of blacks and Hispanics are sampled at a higher rate. Within each secondary sampling unit (SSU) all households containing any eligible black or Hispanic persons are selected for interview, while only a subsample of the other households are selected for interview. Beginning with the 2006 NHIS, households within each sample SSU containing Asian persons also are selected for interview.

The Current Population Survey is a complex multistage stratified sample survey of households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The survey is designed to measure demographic and labor force characteristics of the civilian noninstitutional population 16 years of age and older (Technical Paper 63RV, 2002). The monthly CPS survey weights are ratio adjusted so that estimated population totals from the CPS agree with independent estimates of the civilian noninstitutional population that are produced by the Census Bureau. The NHIS person weights for each calendar quarter are post-stratified using the same independent estimates that the monthly CPS uses for the midpoint month of each quarter (i.e., February, May, August, November).

#### 4. Methods

##### 4.1 Assessing Coverage Distribution of NHIS Sample Adults

With the impending 2006 change in sample selection for NHIS minority persons age 65 years and older, it was important to review the NHIS sample adult distribution over several years (1997-2003) and compare to the CPS population figures. We wanted to see whether the new minority oversampling selection procedures might affect the sample adult coverage distributions.

For each NHIS sample year, we created poststratification cells using three race/ethnicity groups (Hispanic, Black, and White/Other) and six age groups (18-24, 25-34, 35-44, 45-54, 55-64, and 65+). For each cell, we used the nonresponse adjusted sample adult weight to calculate the percentage of persons represented by that cell. This weight does not include a poststratification adjustment. Then, for the corresponding cells, we used fully weighted CPS data to calculate the percentage of persons represented by each cell. The difference between these two percentages was calculated for each poststratification cell, as a simple bias measure. The term bias formally denotes the difference between the expected value of a statistic and its “true” population value. Here we assume that the CPS provides the “true” population value:

$$Bias(y) = E(y) - Y \quad (1.1)$$

The NHIS sample value,  $y$ , is called an unbiased estimate if its expected value coincides exactly with the corresponding true population value  $Y$  (i.e., the bias equals zero). When not zero, the sampling bias is

small in most well-designed samples, and it tends to diminish with increasing sample size (Kish, 1965).

##### 4.2 Expected Effects of Increased Oversampling

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 that contained both elderly and non-elderly (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. When the sample adult’s probability of selection was doubled, the total sample adult effective sample size decreased by an estimated 1.6%, with a corresponding 13.4% increase in yield of minority adults age 65 years and older. 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 picked).

Note: The same demographics were used within each test, but differed between tests.

**Test 1:** 20 repetitions, 3-adult family, 1 female 70+ years old (person #2), 1 sample adult (SA) to be picked.

Person	SA picked	Occurrence
#1	7	35%
#2	10	50%
#3	3	15%
Total reps	20	

With the impending 2006 change in the sample selection procedure, Test 1 was performed in order to simulate the process of oversampling minority sample adults age 65 years and older. After reviewing results from Test 1 above, it was suggested that more repetitions be used to give a better statistical result. Theoretically for Test 1, each regular adult would have a 25% chance of being selected, with the minority 65+ adult (person #2) having a 50% chance. The occurrence of the minority 65+ adult was exactly as expected, while the occurrence of the other two were not, likely because of the small sample that was tested. Results using a larger sample are shown in section 5.2.

### 4.3 Sample Adult Roster Randomization

As part of this research, we examined how the sample adult is selected within the family, and whether some additional form of randomization should be required prior to selection. To select one adult from the family, a procedure for ordering the members of the household is first needed. In the NHIS, all relationships in the household are recorded relative to a household reference person, who is the person who owns or rents the housing unit. If more than one person owns or rents the housing unit, the oldest among them is designated the household reference person; if none of the family members owns or rents the unit, then the oldest person in the family is designated the reference person. For the Family Core, all members of the household 17 years of age and over who are at home at the time of the interview are invited to participate and to respond for themselves. For children under 17 and those adults not at home during the interview or who choose not to participate, information is provided by a knowledgeable adult family member (18 years of age or over) residing in the household. From each family in the NHIS, one sample adult is randomly selected, and information is collected in the Sample Adult Core questionnaire. Prior to 2006, all eligible adults on the roster had the same probability of selection as the sample adult.

In the NHIS, CAPI software is used to select the sample adult from the family roster of adults. Note that family size can vary considerably. Table A shows the distribution of the number of family members.

**Table A. Size of Family, 2003 NHIS**

Number of Members	Percent
1	29.7
2	30.3
3	15.6
4	13.7
5	6.7
6 or more	4.0

In Test 1 above, we saw that person #1 was selected at a higher rate of 35%, rather than the expected 25%. Thus, we wanted to determine how frequently the first person listed on the roster is selected as the sample adult.

## 5. Results

### 5.1 Results of Coverage Distribution Evaluation

Table 1 provides the percent coverage difference results using the 1997 NHIS. The percent coverage difference is obtained by subtracting the fully weighted CPS coverage percentage for the poststratification cell from the NHIS coverage percentage excluding the poststratification adjustment.

**Table 1. 1997 NHIS Percent Coverage Differences from CPS Totals**

Age	Hispanic	Black	Other
18-24	-0.13	-0.49	-1.68
25-34	-0.13	-0.28	-0.15
35-44	-0.17	-0.36	0.38
45-54	0.08	0.06	0.52
55-64	-0.06	0.10	0.55
>=65	<b>-0.08</b>	<b>0.22</b>	<b>1.60</b>

The largest differences appear to occur in the “Other” race category. However the magnitude of none of these differences is greater than 2% indicating the current distribution estimates are similar to population values.

Figure 1 shows the percent differences, for sample adults 65 years and over, between the NHIS nonresponse-adjusted weighted estimates of the total population and the fully weighted CPS estimates for each data year (1997-2003).

### 5.2 Expected Results of Increased Oversampling

In order to provide a better statistical result for Test 1 above, the number of repetitions was increased. For Tests 2 and 4, the simulation included 500 repetitions, and Test 3 incorporated 1000 repetitions. The number of eligible adults in the family was also increased from three to four in Tests 2 and 3 to provide a better picture of the randomness of selection. The demographics for Test 4 included 3 eligible adults with one child.

**Test 2:** 500 repetitions, 4-adult family, one male 80+ years old (person #2), 1 sample adult (SA) to be picked.

Person	SA picked	Occurrence
#1	84	16.8%
#2	215	43.0%
#3	101	20.2%
#4	100	20.0%
Total reps	500	

**Test 3:** 1000 repetitions, 4-adult family, one male 80+ years old (person #3), 1 sample adult (SA) to be picked.

Person	SA picked	Occurrence
#1	192	19.2%
#2	207	20.7%
#3	387	38.7%
#4	214	21.4%
Total reps	1000	

**Test 4:** 500 repetitions, 3-adult family with 5-year-old child (person #4), one female 65+ years old (person #2), 1 sample adult (SA) to be picked.

Person	SA picked	Occurrence
#1	123	24.6%
#2	249	49.8%
#3	128	25.6%
#4	0	0.0%
Total reps	500	

As we expected, with larger samples, Tests 2-4 came very close to the theoretical figures. Theoretically, for Tests 2 and 3 each adult under age 65 would have a 20% chance of selection, and the adult age 65+ years old would have a 40% chance. Theoretically, for Test 4 each adult under age 65 would have a 25% chance of selection and the 65+ years old adult would have a 50% chance. In testing the accuracy of the random number generator, the actual results came very close to these figures indicating that the Random Sample Adult Generator is performing as intended. The observed selection percentages almost always varied from the expected percentages by less than

$$\frac{1}{\sqrt{\text{number of simulations}}},$$

so the minor fluctuations can be attributed to sampling variability. Note that

$$\frac{1}{\sqrt{\text{number of simulations}}} \cong 1.96 \sqrt{\frac{\max[(p)(1-p)]}{\text{number of simulations}}}$$

where the right-hand side of the expression corresponds to the half length of a 95% confidence interval for a proportion (p) equal to 0.5 (50%) with a sample size equal to the number of simulations.

### 5. Conclusion

For the coverage evaluation, the largest difference between the NHIS estimates and the CPS is 2% for

nonminority sample adults. Based on this result, we believe the current distribution estimates are very close to population values and expect to achieve good results with the higher selection probability procedure for minorities age 65 years and older in the NHIS.

The results in section 5.2 also assured us that additional randomization of the sample adult roster would not be necessary. Sample adults were selected in the simulations in the proper proportions, taking account of sampling variability during the simulations.

The next item on the research agenda is to investigate how well the higher selection probability procedure worked in practice, using data from the 2006 NHIS.

### Acknowledgements

We thank Ted Musiker and Curtis Ziesing of the Census Bureau for performing the random selection generator simulations.

### References

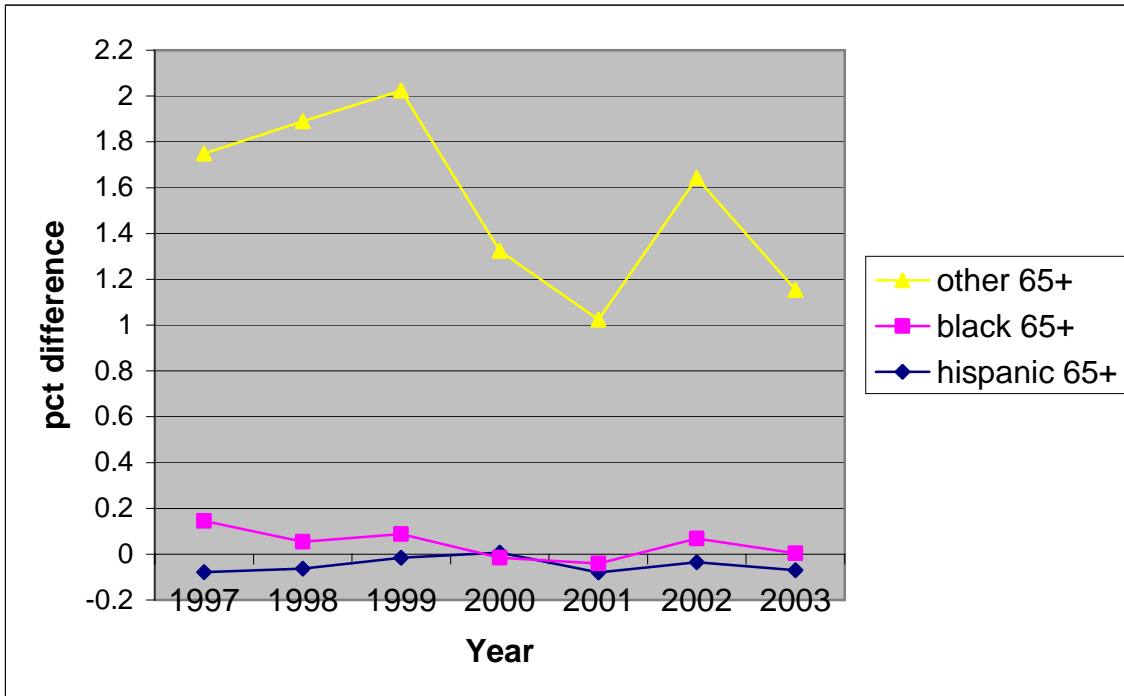
Davis, Karen. 2003. *Two Options for Oversampling in the National Health Interview Survey*. 2003 Proceedings of the Annual Meeting of the American Statistical Association [CD-ROM], Alexandria, VA: American Statistical Association, 1176-1181.

Kish, Leslie, (1965), *Survey Sampling*, New York: John Wiley and Sons.

National Center for Health Statistics (2004). *Data File Documentation, National Health Interview Survey, 2003 (machine readable data file and documentation)*. National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD.

Technical Paper 63RV, *Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2002. <http://www.census.gov/prod/2002pubs/tp63rv.pdf>

**Figure 1. Percent Differences Between NHIS Nonresponse Adjusted Weighted Estimates of the Total Population (not including post-stratification) and Current Population Survey Weighted Estimates of the Total Population (including post-stratification) for Sample Adults 65+ Years Old, by Sample Year and Race/Ethnicity**



Sources: NHIS and CPS 1997 – 2003  
 Percent Difference=NHIS value – CPS value