

INTRODUCTION

A poststratified ratio adjustment is an estimation feature in many household interview sample surveys. This feature modifies a survey estimator so as to yield estimated totals coincident with independent benchmark totals. There are several potential advantages of this estimation feature: (1) It reduces variation between benchmark and estimated survey population totals for major population domains; (2) it reduces the variability of survey estimates (e.g., see [1], [2], and [4]); and (3) it compensates for sampling problems (e.g., problems relating to frame coverage) not addressed by the survey estimator.

Incomplete coverage of the target population by sampling frames for household surveys like the National Health Interview Survey (NHIS) and the Current Population Survey (CPS) may occur for a variety of reasons. The NHIS is a household interview survey with a large annual sample sponsored by the National Center for Health Statistics, and the CPS is a household survey of even larger sample size sponsored by the Bureau of Labor Statistics. Much research has already been targeted on identifying incomplete (and disproportionate) coverage of the population by the CPS frame. This CPS research is also applicable to the NHIS, in part because before 1985 the NHIS sample design was a subset of the CPS sample design; however, the CPS and NHIS differ in other survey features.

According to Brooks and Bailar [3], "the CPS sampling frame does not fully represent the target population. However, frame deficiencies (excluding within household coverage loss) represents less than 3 percent of the population, though they are concentrated. ... Within household misses are believed to account for a larger part of the undercoverage." Brooks and Bailar report "concealment and oversight" as two reasons for respondents to provide incomplete rosters to interviewers with these reasons "applying more to men than to women and more to poor persons than non-poor persons." Moreover, "illegal immigrants have a strong incentive to conceal presence in the household, and persons loosely attached to households may be unintentionally left off rosters by respondents." Also "persons with no attachment to any fixed address present a coverage problem for the CPS."

This paper has two main objectives. Based on the NHIS, this paper illustrates the patterns of the inflation of estimated population totals due to poststratification, including identifying poststrata with relatively large inflations. Also, to identify potential disproportionate coverage of the target population by the NHIS sampling frame, this paper compares, by the NHIS poststrata, selected percentage estimates based on the 1980 NHIS and the 1980 Census sample for selected socio-demographic characteristics. Overall, the NHIS poststratification adjustment inflates the 1984 NHIS estimated U.S. population

total by 10 percent but has differing patterns of inflation by poststrata. These comparisons may understate any NHIS coverage problems resulting from the potential disproportionate coverage of the target population by the NHIS sampling frame, because the sampling frame for the Census sample may also disproportionately cover its target population by race, income, and urban status. Specific coverage problems in the 1980 Decennial Census were cited at an August 17, 1987 Congressional Hearing, conducted in conjunction with these Meetings, before the Subcommittee on Census and Population, Committee on Post Office and Civil Service, on "Census Undercount and the Feasibility of Adjusting the Census Figures."

Before discussing NHIS data one must be familiar with the NHIS survey design. Section 1 sketches the 1973-1984 NHIS sampling, data collection, and estimation plans. Section 2 presents and discusses the inflation due to poststratification in the 1984 NHIS. Since disproportionate survey coverage is speculated to be more pronounced among particular subdomains, Section 3 compares for selected poststrata the estimated percentage for several characteristics of interest, based on the 1980 NHIS and the 1980 Census sample. Section 4 presents the conclusion that there is some evidence of disproportionate frame coverage of the target population.

Note, while NCHS calculates the NHIS poststratified ratio adjustments for each survey calendar quarter, the adjustments presented in this paper are calculated for the annual NHIS (i.e., four calendar quarters).

Section 1. NHIS SAMPLING, DATA COLLECTION, AND ESTIMATION PLANS

The 1973-1984 NHIS uses a questionnaire in a household interview to obtain information on personal and demographic characteristics, illnesses, injuries, impairments, chronic conditions, and other health topics. The NHIS inferential population is the U.S. civilian noninstitutionalized population. While the NHIS design is well documented (see, for example, [1] and [2]), some information on the NHIS survey design is needed in order to discuss the survey's poststratified ratio adjustment.

Sampling plan. The NHIS is based on a multi-stage probability sample. For the 1973-1984 NHIS the U.S. is divided into 1924 primary sampling units (PSU's). In general each PSU is a county, a small group of contiguous counties, or a standard metropolitan statistical area. These PSU's are grouped to form 376 strata, with 156 of these being self-representing. From each stratum one PSU is selected for the NHIS sample. Without a loss of generality, the remaining stages of sampling can be combined and treated as an ultimate stage. The ultimate sampling units (USU's) called segments are defined in such a manner that each segment contains an expected four housing units. Second-stage NHIS sampling rates yield an overall

equal probability sample of USU's. All eligible individuals in the sample housing units are selected for the sample. The usual 1973-1984 annual NHIS sample consists of a probability sample of approximately 41,000 eligible occupied households, yielding about 105,000 persons.

An annual NHIS sample is divided into 52 weekly samples, where each is a national probability sample. Interviewers obtain NHIS data for a sample household for the week specified by the sample assignment.

Data Collection Plan. The Census Bureau is the NCHS data collection agent for the NHIS. Census Bureau interviewers obtained response from 96 percent of the NHIS sample households eligible for the NHIS in 1984.

For each eligible sample household an NHIS interviewer completes a separate questionnaire for each family (individuals related by blood, marriage, or adoption) where socio-demographic and health information is separately obtained for each family member. For NHIS data collection the interviewer attempts to have present all household members 19 years of age or older. When this is not possible, proxy response for adult family members is accepted. Proxy responses are obtained for persons under 19 years of age although persons 18-19 years of age may respond for themselves.

Interviewers are trained and retrained in NHIS data collection materials. As a part of survey documentation the Census Bureau prepares a detailed NHIS data collection manual [5].

Estimation. The NHIS employs a complex estimation procedure. Without a loss of generality the NHIS person sampling weight may be considered the product of four factors. The first factor is the reciprocal of the probability of sample selection (i.e., PSU, segment, and housing unit). The 1973-1984 NHIS is designed to yield an equal probability sample. The second factor is a segment-level adjustment attempting to compensate for household nonresponse. The third factor attempts to reduce the variability of NHIS population estimates resulting from the sampling of nonself-representing PSU's. The weighting classes for this adjustment are race-residence classes in each of the four Census geographic regions. The fourth factor, the poststratified ratio adjustment, aligns NHIS estimated population totals based on the prior three factors to independent estimates of the population, by 60 age, sex and race poststrata. These independent population estimates are based on Census population projections.

Section 2. 1984 NHIS POSTSTRATIFICATION

While the poststratification adjustment inflates the 1984 NHIS estimates for the U.S. population totals by 10 percent, this poststratification disproportionately affects 1984 NHIS estimated population totals by subdomain.

The poststratification adjustment increases the 1984 NHIS estimated population total for Blacks by 20 percent. This adjustment also increases the NHIS estimated population total for nonBlacks by 9 percent.

For Blacks the poststratification inflates the NHIS estimated population total for males (23 percent) more than it inflates the population total for females (17 percent). Among nonBlacks the adjustment similarly increases NHIS estimated population totals for males (9 percent) and for females (9 percent).

Table 1 presents the poststratified ratio adjustments for the 1984 NHIS. For example, the poststratified ratio adjustment increases the NHIS estimated population total for Black males 25-29 years of age by 54 percent, because the ratio adjustment is 1.54. The 1984 poststratified ratio adjustments for the other poststrata are smaller. Since persons of the White race dominate the category for nonBlacks, estimates for characteristics of interest for nonBlack persons to Whites. Some pronounced patterns of poststratification inflation by age among poststrata are observed too. Among individuals 17 years of age or older the largest poststratified ratio adjustment by age are among those 17-44 years of age and among those 55 years of age and older.

Based on table 1, some differences by age and race are observed. Generally the largest poststratified ratio adjustments are those for Black males, and the smallest ratio adjustments are those for nonBlack females. This level of poststratification inflation, coupled with large poststrata sample sizes, motivated this investigation on whether there is evidence of disproportionate population coverage by the frame according to education and marital status.

NCHS generally does not produce NHIS statistics with age categories as fine as those in the poststrata. One set of response categories for age in years used for NHIS publications is <6, 6-16, 17-44, and 45 or over [6]. This collapsing of poststrata should increase the stability of estimates. Still relatively large inflation factors due to poststratification occur throughout age intervals covering individuals 20-49 years of age. We later look closer at data for poststrata covering these age intervals. Since the 1984 NHIS contains data on over 100,000 sample persons, most NHIS percentage and aggregate estimates for major subdomains (including poststrata) are based on large samples.

Several possible explanations exist on the causes of such patterns of poststratification inflation. As noted, Brooks and Bailar detail various (coverage and non-coverage) deficiencies in the sampling frame for the CPS that may also apply to the NHIS. Still only limited qualitative and quantitative information seems to be available on specific population subdomains with potential coverage deficiencies.

The 1980 Census sample [7] yields estimated percentage distributions of population by selected characteristics of interest according to NHIS poststrata. Percentage estimates for major population domains from the 1980 Census sample are subject to negligible sampling error, because of the size of the Census sample and its sampling rate. Approximately 1 out of 5 households in Nation were selected for the 1980 Census sample.

Still, because of potential disproportionate coverage of the population the Census sample may

underestimate population totals for some domains. Some of these domains may overlap the domains where there is potential disproportionate coverage of the target population by the NHIS sampling frame. For example, "In 1980, the Census Bureau believes that [for the Decennial Census] it failed to physically find about 5.6 percent of the Nation's black citizens, at least the same share of Hispanics, and 0.5 percent of all whites [August 4, 1987, page 37, "Wall Street Journal"]".

In this paper the 1980 Census sample data are compared with that data from the 1980 NHIS, not that from the 1984 NHIS. This allows estimation of characteristics of interest from the NHIS and the Census sample for a more common reference period.

Table 2 displays the distribution of response to the 1980 NHIS according to poststrata. The poststratum with the smallest sample has 109 sample persons. On the average each poststratum contains about 1700 NHIS sample persons.

Note the overall inflation effect of poststratification in the 1980 NHIS is smaller than that for 1984 NHIS--the overall inflation was about 10 percent in 1984 and 5 percent in 1980. The 1980-1984 increase in the overall inflation due to poststratification may be in part due to the aging of the NHIS sample design, which decreases its efficiency. Even so, in the next section some evidence is provided for disproportionate NHIS coverage.

Section 3 COMPARING ESTIMATES FROM 1980 NHIS AND 1980 CENSUS SAMPLE

This section displays selected percentage estimates for selected characteristics of interest based on the 1980 Census sample and the 1980 NHIS for several NHIS poststrata. These comparisons should be, however, interpreted with caution.

These survey's inferential populations, procedures, reference periods, and definitions differ. For example, active duty military personnel and the institutionalized are in the inferential population for the 1980 Census sample but not in that for the NHIS. In 1980 the civilian noninstitutionalized U.S. population--the NHIS inferential population--is 96 percent of the resident population. The 1980 Census sample is based on data collected as of Census Day (April 1) in 1980, while the 1980 NHIS is based on data collected throughout 1980.

As noted it may be more difficult to obtain survey data from persons in lower income households and from persons with less education. Concealment and oversight are the two reasons previously cited as causing reporting problems for within household listings. However since interviewers for the NHIS and CPS routinely probe for any eligible unlisted persons on a household roster, "concealment" appears a more likely reason for nonlisting than "oversight." Lower response rates may also occur in certain geographic regions and in some large urban areas. We examine the estimated proportion of adults by age, race, and sex (NHIS poststrata) according to educational attainment, marital status, and household size. Individuals having

less than a high school education, for example, may contain a disproportionate number of individuals with lower health status.

By comparing 1980 Census sample and NHIS percentage estimates we found some evidence of disproportionate NHIS undercoverage by education and by marital status. In particular we are interested in whether the NHIS estimates of the proportion of the population with less than a high school education are less than the figures for the corresponding domain from the 1980 Census sample.

Table 3 displays the estimated proportions of persons 20-49 years of age based on the 1980 Census sample and the NHIS by completion of fewer than 12 years of education according to poststrata. Some pronounced patterns occur.

The 1980 Census sample yields a significantly larger proportion of males 15-44 years of age with fewer than 12 years of education. This relationship is also true within poststrata for males 20-49 regardless of race and age. Although few of these differences by poststrata are significant and some differences are quite small, this pattern of differences appears unlikely to be attributable to chance. One explanation is that there is disproportionate coverage of males with less than a high school education in the target population by the sampling frame. Among the poststrata for women there is no pattern between estimates based on the NHIS and the Census occurs.

In addition table 3 displays similar data for the estimated proportion of individuals who were never married. The 1980 Census sample yields a significantly larger proportion of never married males 15-44 years of age. Table 3 shows that this relationship is also true within poststrata for males 20-44 years of age regardless of race and age. Since this suggests that the differences may be related to family size, we compare Census sample and NHIS estimated proportions of the population living alone by stratum.

Table 3 displays the data for the estimated proportion of the population living alone by age and sex based on the 1980 Census sample and the NHIS. Since proxy response is not possible in one-person households, the Census sample and the NHIS may be less likely to obtain a response from single-person households. Since the prior patterns were not replicated for this variables, the potential coverage problem for those with less than a high school education and for those never married does not seem to carry over for individuals living alone. This result is consistent with the Brooks and Bailar report that within-household coverage problems are more severe than the coverage problems of households themselves.

Such difference are important because they may suggest that the NHIS sampling frame has a disproportionately lower coverage rate of the target population among certain subdomains whose members may have poorer health status.

Table 3 present data on these characteristics of interest based on a self-weighted NHIS estimator. This self-weighted estimator basically is the final NHIS estimator without first-stage and second-stage ratio adjustments.

With these additional estimates we can isolate the approximate effect of the NHIS first-stage and second-stage ratio adjustments.

These adjustments do not seem to cause the NHIS estimates to depart from those from the Census sample. The first-stage adjustment is not expected to have a large effect on estimates related to minorities and poor, since the adjustment does not affect data from the Nation's largest urban areas, where a large proportion of the Nation's poor and minorities reside.

Section 4. CONCLUSION

A comparison by poststrata of 1980 NHIS and 1980 Census sample data for the proportion of the population 20-49 years of age with less than a high school education or having never married showed a potentially significant but relatively small difference for males. This provides some evidence on disproportionate coverage of the NHIS target population by the sampling frame. Within age and race categories the actual difference varied--although differences by race and age categories were not statistically significant. Based on this result it appears unlikely that disproportionate coverage of the population by the sampling frame according to educational status or marital status would substantially affect NHIS estimates other than potentially for the cited subdomains.

References:

- [1] National Center for Health Statistics, M. G. Kovar and G. S. Poe: The National Health Interview Survey Design, 1973-1984, and Procedures, 1975-1983. Series 1, No. 18, DHHS Pub No. (PHS) 85-320.
- [2] National Center for Health Statistics, J. A. Bean: Estimate and Sampling Variance in the National Health Interview Survey, Series 2, No. 38, DHEW Pub. No. (HRA) 74-1288.
- [3] Office of Federal Statistical Policy and Standards, C. A. Brooks and B. A. Bailar: Statistical Policy Working Paper 3, An Error Profile: Employment as Measured by the Current Population Survey, September 1978.
- [4] Bureau of the Census, Technical Paper 40, The Current Population Survey, Design and Methodology, September 1978.
- [5] Bureau of the Census, The 1983 National Health Interview Survey Interviewer's Manual (HIS-100), periodically revised, unpublished document.
- [6] National Center for Health Statistics, S. B. Jack: Current Estimates from the National Health Interview Survey, U.S., 1980, Series 10, No. 139, DHHS Pub. No. (PHS) 82-1567.
- [7] Bureau of the Census, 1980 Census of the Population, Volume 1: Characteristics of the Population, Chapter D: Detailed Population Characteristics, Part 1: United States Summary, Section A: United States Tables 253-310.

Table 1. Inflation Factor Due to NHIS Poststratified Ratio Adjustment by Poststrata. 1984.

Race and Age	Sex	
	Male	Female
	<u>Inflation Factor</u>	
<u>Black</u>		
Under 1	1.42	1.49
1-4	1.26	1.28
5-9	1.18	1.08
10-14	0.99	1.14
15-16	1.09	1.13
17-19	1.20	1.07
20-24	1.37	1.33
25-29	1.54	1.20
30-34	1.33	1.17
35-44	1.27	1.11
45-49	1.20	1.08
50-54	1.07	1.08
55-64	1.24	1.20
65-74	1.20	1.16
75+	1.30	1.25
<u>NonBlack</u>		
Under 1	0.99	1.05
1-4	1.01	1.06
5-9	1.04	1.08
10-14	1.06	1.02
15-16	0.95	1.04
17-19	1.10	1.14
20-24	1.19	1.14
25-29	1.15	1.09
30-34	1.11	1.10
35-44	1.12	1.07
45-49	1.08	1.04
50-54	1.01	1.06
55-64	1.14	1.12
65-74	1.09	1.12
75+	1.10	1.11

Table 2. Unweighted Number and Distribution of the 1980 NHIS Sample By Poststrata

Race and age	Sex	
	Male	Female
	<u>NHIS Sample</u>	
<u>Black</u>		
Under 1	112	126
1-4	493	419
5-9	545	572
10-14	577	623
15-16	273	255
17-19	373	372
20-24	449	611
25-29	419	559
30-34	357	484
35-44	434	638
45-49	210	305
50-54	208	306
55-64	363	517
65-74	258	355
75+	109	190
<u>NonBlack</u>		
Under 1	700	665
1-4	2,656	2,552
5-9	3,397	3,363
10-14	3,729	3,571
15-16	1,709	1,562
17-19	2,334	2,457
20-24	3,854	4,015
25-29	3,620	3,954
30-34	3,459	3,704
35-44	5,181	5,468
45-49	2,243	2,405
50-54	2,332	2,532
55-64	4,369	4,944
65-74	2,915	3,716
75+	1,373	2,281

Table 3. Percentage of the Population 20-49 With Less Than a High School Education, Never Married, and Living Alone, by Race and Age (NHIS postrata). 1980 NHIS and 1980 Census Sample

Race and Selected Age Groups	Sex					
	Male			Female		
	Census Sample Estimate	NHIS Self-Weighted Estimate	NHIS Final Estimate	Census Sample Estimate	NHIS Self-Weighted Estimate	NHIS Final Estimate
<u>Percent with less than a high school education</u>						
Black						
20-24	30.2	29.6	29.8	23.3	23.0	23.0
25-29	26.2	23.1	23.2	23.5	19.5	20.2
30-34	27.6	21.6	21.5	26.6	28.2	28.3
35-44	39.5	38.4	39.1	33.9	36.2	36.4
44-49	52.1	44.7	42.9	49.6	46.3	46.2
NonBlack						
20-24	17.7	17.1	17.1	15.3	16.5	16.5
25-29	14.3	14.0	13.9	14.1	14.4	14.4
30-34	14.3	13.6	13.8	15.0	14.7	14.2
35-44	21.1	21.0	21.0	21.9	21.4	21.1
45-49	28.4	25.8	25.4	28.6	27.0	26.8
<u>Percent never married</u>						
Black						
20-24	77.8	75.9	75.9	67.5	60.9	60.7
25-29	42.6	39.4	38.9	37.0	35.4	34.9
30-34	23.1	17.9	18.3	21.5	20.0	20.7
35-44	13.5	12.4	12.6	12.7	13.8	12.7
44-49	10.1	10.1	8.7	8.4	9.8	10.1
NonBlack						
20-24	66.9	62.8	62.0	48.6	43.8	43.3
25-29	30.8	26.5	27.0	19.3	16.8	16.7
30-34	14.0	13.0	13.1	9.2	8.7	9.0
35-44	7.2	6.1	6.2	5.2	5.2	5.4
45-49	5.5	5.4	5.2	4.2	4.4	4.6
<u>Percent living alone</u>						
Black						
20-24	8.0	15.6	13.6	5.2	6.9	5.9
25-29	12.9	13.6	13.3	7.2	6.1	5.9
30-34	12.9	12.6	12.0	6.1	6.8	6.8
35-44	12.2	10.5	10.3	5.9	8.5	8.5
44-49	12.9	13.5	12.7	8.5	11.1	10.8
NonBlack						
20-24	9.1	10.5	7.5	5.9	9.0	7.1
25-29	12.3	12.3	12.6	7.3	7.9	7.7
30-34	10.0	10.2	10.0	5.3	6.7	6.9
35-44	7.3	6.7	6.8	3.7	5.1	5.0
45-49	6.6	7.7	7.9	5.2	6.1	6.1