VARIABILITY IN THE SAMPLING WEIGHTS IN THE NATIONAL HEALTH INTERVIEW SURVEY--CAUSES, IMPLICATIONS, AND STRATEGIES

Steven L. Botman
NCHS, 6525 Belcrest Rd., Hyattsville, MD 20782

KEY WORDS: survey design, variances, and mean square error

1 Introduction

The National Health Interview Survey (NHIS) is a multipurpose national survey, sponsored by the National Center for Health Statistics, based on a stratified multistage area probability sample.[1] NHIS data sets are based on the NHIS Basic Health and Demographic questionnaire or a questionnaire for an NHIS supplemental health topic. This paper, however, is limited to the distribution of the final and other annual sampling weights in the 1991 NHIS data set and discusses the cause of variability in these weights and some implications. Some potentially reasonable re-weighting strategies are also discussed that reduce such variability.

```
15500+        *
     ]
     ]        *
12500+        *
     ]        *
     ]
 9500+        *
     ]        *
     ]        *
 6500+        *
     ]        *
     ]        *
 3500+        0
     ]     +--0--+
     ]     *--0--*
  500+        0
```

Figure 1. Boxplot for Final Annual Weights

These alternative weighting strategies reduce the largest weights for survey observations, which for several reasons are potentially the most troublesome extreme weights. The variability of the final annual NHIS sampling weights is related to several features in the NHIS sample design. This paper gauges the variability in these weights and explores their effect on the values, variances, and mean square error of resultant estimates.

Figure 1 show a boxplot from SAS[2] for the final sampling weights for the NHIS sample. While the final NHIS sampling weights for the 120,000 persons in the sample are clustered near the average value of about 2,000, the boxplot shows the distribution of these sampling weights. For only four persons in the sample is the final annual sampling weight greater than 10,000, which is about 5 times the average value of these weights. For about 100 persons, the final sampling weights are less than 10,000 but above 8,500.

In addition to other implications, this plot clearly demonstrates that some observations in the sample have a larger influence on the survey estimates than others. Moreover, this plot clearly demonstrates that some observations in the sample have a substantially less influence on the survey estimates than others.

Section 2 discusses how features of the NHIS sampling plan are related to the distribution of the NHIS sampling weights. Section 3 discusses how features of the NHIS weighting procedures are related to the distribution of the sampling weights. Section 4 provides some quantification on the distribution of these weights for domains defined by race, sex, and age. Section 5 discusses some approaches to reduce such variation in the weights. Section 6 summarizes the paper.

2 Causes of a Variability of Sampling Weights--NHIS Sampling Plan

To increase the statistical reliability of the NHIS estimates for Black Americans, the NHIS oversamples persons residing in areas with high concentrations of Black Americans.[1] In the NHIS area sample in primary sampling units (PSUs) with between 5-50 percent population of Black Americans, second-stage sampling units are partitioned into two substrata based on the concentration of Black Americans. In the area substrata, second-stage units are sampled at a higher rates in the high concentration Black stratum and are sampled at a lower rate in the other stratum. In many PSUs, a stratum for housing units selected from a list of building permits is also used.

Figure 2 shows a boxplot for the final annual NHIS sampling weights from SAS by the three (sub) strata used within PSUs. The sampling plan is clearly a cause of the variability of the sampling weights. The largest variation is among the final annual weights in the substrata for areas not oversampled for Blacks.

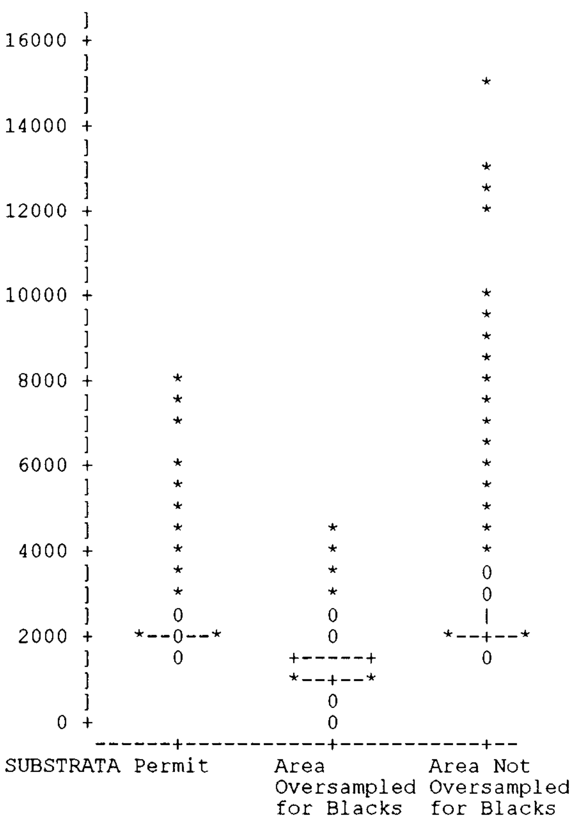

Figure 2. Boxplot for Annual Final Weights by Substrata

Figure 2 shows that the average sampling weight are about 1,900 for the permit (sub) stratum, 1,100 for the (sub) stratum of the areas oversampled for Black Americans, and 2,000 for the (sub) stratum of the areas not oversampled for Black Americans. The ratio of the average weights in the oversampled areas to those in non-oversampled areas is about 2 to 1.

In addition, the NHIS subsamples some designated second-stage units in areas where interviews must travel unusually far distances to reach the PSU. Typically this includes some second-stage units in Alaska and Hawaii. Interviews in these areas, thus, are conducted by interviewers traveling from the contiguous U.S. and thus are quite costly.

Moreover, the NHIS is a based on an area sample. Sometimes when interviewers list households in second-stage sampling units, they find substantially more households than expected. In such cases to maintain reasonable interviewer workloads, second-stage units are subsampled for survey data collection. The use of a building permit stratum when available for new construction generally avoids this problem, because it provides a frame of units within the PSU built since the last Decennial Census. Still not all areas of the country are covered by offices issuing such permits.

3 Causes of a Variability of the Sampling Weights--Weighting Procedures

Each year, the NHIS introduces a new set of supplemental questionnaires. To train interviewers, NCHS from 1987 through 1992 has canceled interviewing for the sample assigned to the first week in the survey year. While the NHIS general estimation strategy is described elsewhere,[1] this general strategy is modified by doubling the "base" weights for the sample assigned to the second week of the survey year when the sample assigned to week 1 is dropped. This is employed as bias reduction strategy to better reflect the number and extent of acute conditions (e.g., influenza) usually found in the early January in the population. The variability of such results will, for example, depend on the severity and extent of any influenza outbreaks that year. Another weighting approach, for example, would be to uniformly increase the base weights for the first quarter by 13/12 when the sample for the first week of that quarter is dropped.

Table 1 illustrates the situation for the first

Table 1 Estimated Average Number of Restricted Activity Days (AV_RAD), Average Number of Bed Days (AV_BED), Average Number of Doctor Visits (AV_TDV) in the Two Weeks Prior to the Week of Interview per Thousand Persons, Quarter 1, 1991 NHIS (Base Weights)

Week	AV_RAD	AV_BED	AV_DV
2	751	303	219
3	689	274	238
4	734	327	223
5	579	274	223
6	759	311	226
7	629	293	198
8	711	330	222
9	697	277	228
10	697	288	233
11	637	282	206
12	576	236	260
13	574	202	209

```
       ]
16000 +
       ]
       ]                        *
       ]
14000 +
       ]
       ]                        *
       ]                        *
12000 +                         *
       ]
       ]
       ]
10000 +          *
       ]          *
       ]          *             *
       ]          *             *
 8000 +          *             *
       ]          *             *
       ]          *             0
       ]          *             0
 6000 +          *             0
       ]          *             |
       ]          *             |
       ]          *          +-----+
 4000 +          *          *--+--*
       ]          0          +-----+
       ]          0             |
       ]          0             0
 2000 +       *--+--*          0
       ]          0             0
       ]          0             0
       ]          0             *
    0 +          *
      -----------+-------------+----------
 WEEK          Week 3-52     Week 2
```

Figure 3. Parallel Boxplots for Final Annual Weights by Weekly Sample Assignment

quarter of 1991 where, for example, the estimated average number of restricted activity days is appears higher for the first three weeks of the quarter than the last three weeks. There also, for example, are strong seasonal patterns on mortality from influenza.[4]

Figure 3 shows that the largest weights are in the sample assigned to week 2 of the survey year. These large weights are directly attributable to double weighting the base weights for the sample in the week 2, because the sample from week 1 was dropped from data collection for interviewer training. In addition, the NHIS calibrates the survey estimators to independent estimates of the population. This calibration (like other post-survey weighting adjustments) introduces some additional variation in the weights.

4 Distribution of the NHIS Sampling Weights for Race, Sex, and Age Domains

Table 2 shows that overall the coefficient of variation (CV) of the sampling weights is 0.34. The CV of the sampling weights for Black persons is larger than that for nonBlack person. The CV of the weights for some of the domains of Black persons by age and race are much higher than that for similar domains of nonBlack persons.

Table 2 shows the ratio of the largest weight to the smallest weights in these same categories. Overall some cases have 71 times larger sampling weights than others. In the race-sex-age categories this ratio is often over 20 to 1. This table demonstrates that some observation will have a substantially larger effect on survey estimates than the effect of others.

The possibility of influential observations is magnified for characteristics other than attributes. For example, if in a subdomain one of these cases with a relatively large sampling weight had an unusually large number of doctor visits or hospital visits, the one observation could account for a large proportion of a national (subdomain) estimate.

5 Some Approaches

If one analyses NHIS data using variables not related to acute conditions, one might want to consider the beneficial effect on sampling variances of reweighting the file without double weighting the base weights in the sample assigned to the second week of the survey (when the sample for week 1 is dropped). Of course, one would have to well document the approach in order to one's data reproducible and the provide a cogent rationale for the particular way that one modifies the weights.

Disproportionate sampling can be used as a variance reduction technique in an optimized sampling design, where disproportionately fewer units are sampled in strata with high interviewing costs and disproportionately more units are sampled in strata with higher population variance. In other circumstances, disproportionate sampling decreases survey efficiency by increasing the design effect of survey estimator.

Kish[4] notes in these cases that, the increase in the design effect due to disproportionate sampling can be gauged by the relative variance (the square of coefficient of variation) of the sampling weights.

Such reweighting cuts the ratio of the largest sampling weight to the smallest sampling weight overall and for each race-sex-age subdomain. This approach, however, modestly reduces the CV of the sampling weights, cutting the overall CV of the weights by about 3 percentage points.

Another approach might be to also truncate the larger weights. The NHIS where the base weights for week two were not doubled was then reweighted with the largest weights prior to the age sex race adjustment truncated for Blacks at 4101 and for nonBlacks at 3606. These values are about 2 times the average weights for nonBlacks and 3 times the average weight for Blacks and the truncation affects about 1 percent of the cases. This approach cut the ratio of the weights prior to poststratification from 31 to 1 to 19 to 1.

The effect of these weighting strategies on survey estimates and variances may depend on the particular characteristic of interest and domains of study. To obtain clues on the effects we look at a variable strongly related to the time of the year.

The estimated number of restricted activity days within two weeks prior to the date of interview varies by week over the first quarter. Here we expect mainly bias effects, with the estimate obtained under the current weighting scheme presumed as closer to the truth. The use of base weights where the base initial from week 2 are not doubled or the use of truncated weights appear to reduce the sampling error, although any changes in sampling error are quite small.

Table 3 shows the affect of these strategies on aggregate estimates and sampling variances. In general, sampling variances are reduced as one moves from the current NHIS weighting scheme to a scheme where one does not double the base weights for the sample in the second week to one where one truncates larger base weights. For the 1991 NHIS, the changes in the estimates and variances are generally quite small.

Thus the penalty on variances for doubling weights from the NHIS sample for the second week is small on statistics that are not affected by seasonality. Using a mean square error argument and considering the data obtaining using the current weights as truth, we find for the 1991 data some evidence for variables related to early winter health problems in the population that the current strategy of doubling the base weights for the second week sample slightly reduces the mean square error over that obtained using base weights where those from week 2 are not doubled.

6 Summary

The variability in the NHIS weights is related to a number of sampling and estimation features. In addition to implications on sampling variances, one clear implication is that some observation in the NHIS are much more influential than others. Modification of the weights reduces the effect of any unusual observation with a large sampling weight. Cases with large weights still will be more influential on survey estimates. Because this can be especially problematic in subdomain analysis, analyst should review the distribution of weights.

While some of the variability in the weights can be reduced by potentially reweighting, the effect on the variances will likely be small.

References

1. Massey, J. T., Moore, T. F. , Parsons, V. L., and Tadros, W. Design and Estimation for the National Health Interview Survey, 1985-94

National Center for Health Statistics, *Vital and Health Statistics*, 2(110), 1989.

2. SAS Institute, Cary, NC.

3. CDC. Update Influenza Activity--United States, 1992-1993 Season, *Morbidity and Mortality Weekly Report*, 1993; 42:385-387)

4. Kish, Leslie, Weighting for Unequal P_i, *Journal of Official Statistics*, Volume 8, Number 2, 1992.

Table 2: Raw Sample Size (n) and Coefficient of Variation (CV), CV^2, and Ratio of Largest to Smallest for the 1991 NHIS Final Sampling Weights, by Age, Race, and Sex

Race	Sex	Age	n	CV	CV^2	Ratio
.	.	.	120,032	0.34	0.12	71
.	Female	.	57,324	0.34	0.11	61
.	Male	.	62,708	0.34	0.12	70
NonBlack	.	.	102,193	0.31	0.09	61
Black	.	.	17,839	0.50	0.25	69
NonBlack	Female	.	49,450	0.31	0.09	61
NonBlack	Male	.	52,743	0.31	0.09	55
Black	Female	.	7,874	0.51	0.26	55
Black	Male	.	9,965	0.49	0.24	69
NonBlack	Female	0-4	4,071	0.31	0.10	26
NonBlack	Female	5-17	10,141	0.31	0.10	42
NonBlack	Female	18-24	4,501	0.33	0.11	23
NonBlack	Female	25-44	15,816	0.30	0.09	40
NonBlack	Female	45-64	9,557	0.29	0.09	35
NonBlack	Female	65-69	1,876	0.28	0.08	21
NonBlack	Female	70-74	1,608	0.28	0.08	21
NonBlack	Female	75+	1,880	0.30	0.09	21
NonBlack	Male	0-4	3,926	0.31	0.10	23
NonBlack	Male	5-17	9,313	0.30	0.09	41
NonBlack	Male	18-24	4,678	0.33	0.11	21
NonBlack	Male	25-44	16,768	0.30	0.09	42
NonBlack	Male	45-64	10,373	0.29	0.09	32
NonBlack	Male	65-69	2,354	0.29	0.09	21
NonBlack	Male	70-74	2,036	0.31	0.10	25
NonBlack	Male	75+	3,295	0.33	0.11	24
Black	Female	0-4	932	0.50	0.25	35
Black	Female	5-17	2,191	0.48	0.23	39
Black	Female	18-24	753	0.49	0.24	29
Black	Female	25-44	2,028	0.49	0.24	27
Black	Female	45-64	1,291	0.47	0.23	28
Black	Female	65-69	253	0.43	0.18	12
Black	Female	70-74	208	0.65	0.42	37
Black	Female	75+	218	0.56	0.32	23
Black	Male	0-4	937	0.49	0.24	19
Black	Male	5-17	2,157	0.50	0.25	44
Black	Male	18-24	1,023	0.48	0.23	21
Black	Male	25-44	3,074	0.47	0.22	29
Black	Male	45-64	1,786	0.45	0.20	26
Black	Male	65-69	315	0.66	0.44	39
Black	Male	70-74	282	0.54	0.29	19
Black	Male	75+	391	0.47	0.22	13

Table 3. Estimated Number of Two Week Restricted Activity Days and Standard Errors Under 3 Potential NHIS Weighting Schemes

Sex and Race	Poststratified Usual Base Weights		Poststratified Usual Base Weights Without Doubling Those from Week 2		Poststratified Mildly-Truncated Base Weights	
	x	SE(x)	x	SE(x)	x	SE(x)
Total	153,617,192	2,767,713	153,258,286	2,772,373	152,812,858	2,734,615
Black	20,864,851	746,771	20,865,252	740,052	20,803,830	726,660
NonBlack	132,752,340	2,601,364	132,393,034	2,604,273	132,009,027	2,569,343
Males	65,388,174	1,401,220	65,406,964	1,407,051	65,239,806	1,375,789
Black	8,164,136	440,377	8,224,504	442,927	8,218,837	440,943
NonBlack	57,224,038	1,325,817	57,182,460	1,326,351	57,020,969	1,294,648
Females	88,229,018	1,883,214	87,851,322	1,870,458	87,573,052	1,833,737
Black	12,700,715	526,904	12,640,748	512,427	12,584,994	496,813
NonBlack	75,528,302	1,757,568	75,210,574	1,746,068	74,988,059	1,717,594