Application of a Fay-Modified Balanced Repeated Replication Method to the National Health Interview Survey (NHIS)[1]

Van L. Parsons
National Center for Health Statistics, 3311 Toledo Rd, Hyattsville, MD 20782

Abstract

The NHIS is a complex survey targeting the health of the U.S. population. To allow for design-based standard errors via linearization, the public-use data provides limited design structures: strata, PSUs and a final survey weight. These structures are accommodated very well by popular complex-survey data analysis software. However, the linearization approach imposes limitations on statistical functional forms, and furthermore, this approach often does not account for statistical variability due to weighting adjustments. The Fay-modified balanced repeated replication (Fay-BRR) method provides a flexible alternative to variance estimation which can correct for some of the linearization method's weaknesses. In this paper the Fay-BRR method is implemented on the NHIS adult samples for the 1997-2005 NHIS design years, and the operating characteristics are discussed.

Key Words: variance estimation

1. Introduction

The NHIS is based on a complex-survey design that collects data to measure the health characteristics of the U.S. civilian non-institutionalized population. Annual NHIS public-use micro-level data files are released containing survey-design structures which allow for design-based estimation. Currently, the design components available on the public-use files consist of survey weights, strata, and primary sampling units (PSUs) from which the analyst can compute direct survey estimates and their standard errors.

These fundamental design structures are accommodated very well by popular survey analytical software, e.g., R, SUDAAN, SAS and Stata. These popular packages can directly use the provided survey weights, strata and PSUs to evaluate many commonly used statistics, e.g., totals, means, ratio of totals and regressions, along with corresponding standard errors estimated by pre-programmed Taylor-linearization methods.

While the use of the Taylor-linearization is an acceptable approach to variance estimation, the method does have limitations. First, a statistic of interest may have a complicated functional form for which linearization is difficult to implement, or possibly

[1]*The findings and conclusions in this report are those of the author and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.*

no direct linearization exists as in the case of non-smooth estimators. In particular, most software avoids any attempt to linearize the weighting adjustments that lead to the final survey weight. For example, if a unit's final survey weight has a poststratification adjustment, it will introduce a component of sampling variability to the targeted estimator; most pre-programmed Taylor-linearization-based algorithms assume that this weight can be treated as a non-variable sampling weight.

Estimation using linearization-based standard errors for non-smooth statistics can be problematic. Indirect methods involving projections of linearization-based confidence intervals of the cumulative distribution function have been developed for estimating standard errors of percentiles, but when considering other non-smooth statistics, e.g., a constrained regression, a linearization solution may not exist.

The *Fay-modified balanced repeated replication* method (Fay-BRR) (Judkins, 1990 and Rao and Shao, 1999) provides a flexible alternate to linearization for variance estimation. In this paper is a discussion of some of the issues involved in using the Fay-BRR method over the NHIS design cycle from 1997-2005. By example, the impact of the choice of the specific Fay-factor adjustment intrinsic to this method is demonstrated for estimating standard errors for estimated means, regressions and percentiles for "large" and "small" NHIS sample sizes.

2. The NHIS Survey Design

Documentation for the NHIS covering the years 1997 to 2005 is available from the NCHS webpage http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm, and the design structure is detailed in Botman et al. (2001). Some important design features are

1. The NHIS samples about 30K-40K households per year (90K-100K persons).

2. The U.S. geography is hierarchically clustered by state strata, universe PSUs (county or county clusters) and Census-defined block-clusters.

3. Hispanic and non-Hispanic black minority groups are oversampled with respect to population representation.

4. Base sampling weights are adjusted every quarter for household-level non-response and person-level weights are poststratified to Census population control totals.

5. The true sampling design does not have an easily expressed stochastic form. For both variance estimation and geographical disclosure avoidance reasons, some original sampling clusters may have been mixed or collapsed on the public-use files. For practical design-based analyses the 1997-2005 NHIS design will be treated as having 339 strata with two PSUs sampled with replacement from each stratum. These structural strata and PSUs are consistently defined in the public-use databases for 1997 to 2005.

3. Replication Methods for Complex Surveys

As alternatives to the linearization approach to variance estimation are the replication methods. Three commonly used methods are the *bootstrap, jackknife* and *balanced repeated replication* (*BRR*) procedures. The reader is referred to Wolter (2007) for detailed discussion on theory and implementation along with pros and cons of these methods. When these methods are chosen by data producers for usage with public-use micro-data, a set of replicate weights accompanies the released public data. The popular statistical packages mentioned earlier have options for using these replicate methods. Statistics Canada provides bootstrap weights for many of its surveys, the California Health Interview Survey provides jackknife weights, and the NHANES III survey provides Fay-BRR weights.

3.1 Fay-Modified BRR

The *Fay-Modified BRR* approach (Fay-BRR) to complex-survey variance estimation will be evaluated for the NHIS. Rao and Shao (1999) provide discussion on theory and practical implementation. Stated briefly, if a sampling design can be represented as a "2 sampled PSUs per stratum" design, and $wt^{(0)}_{hki}$, $wt^{(0)}_{hk'j}$ are original sampling weights corresponding to two units, *i* and *j*, within paired PSUs, *k* and *k'*, within stratum, *h*, then the Fay-BRR replicate weights, indexed by *r*, are of the form:

$$wt^{(r)}_{hki} = (2-f) \bullet wt^{(0)}_{hki} \text{ and } wt^{(r)}_{hk'j} = f \bullet wt^{(0)}_{hk'j}$$

for a given *Fay-factor*, *f*, in [0, 1), with the within-strata orderings of the (2-*f*) and *f* determined by a Hadamard matrix replication component. (Details are in Rao and Shao, 1999.) If $f = 0$, then these replicates are the standard BRR replicates.

The main reasons for selecting this replication approach in the current study were:

1. As a replicate method, the BRR procedure can be used to define a reproducible set of replicates as a function of a specific Hadamard matrix. The BRR method easily accommodates the "2 sampled PSU per stratum" design imposed on the NHIS. The BRR method is highly structured, unlike the bootstrap whose replicates are based on a random selection algorithm. The jackknife replicates perform poorly for estimating variances for non-smooth estimators, e.g., medians, and could not be recommended as a single global procedure for NHIS variance estimation.

2. BRR replicates can be used to estimate variances for a wide variety of statistics. The Fay-factor provides a tuning parameter which allows flexibility for defining reasonable performing replicates for both smooth and non-smooth statistics.

3. The NHIS weighting adjustments can be incorporated into the replicate weights, thus the construction of final survey weights can be treated as a component of the variability. For the NHIS the non-response adjustment is always defined within a PSU, so the full sample and replicate sample non-response adjustments are identical. The post-stratification adjustment will vary by replicate and thus reflect variability. The Fay-factor reduces the impact of small or empty poststratification cells that may result from a traditional BRR implementation.

4. If given a set of replicate weights (BRR, Bootstrap or Jackknife), one can compute variance estimates for complex survey statistics without resorting to specialized software.

The main disadvantages of using replicates with the NHIS data are hardware and software related. Using NHIS replicate weights over multiple years of sample may require gigabytes of disk storage space. For example, about 80MB storage units are needed per 100 replicates for each 100,000 sample units, thus a multiple year analysis, say with 300 replicates and one million records would require 2.4GB storage units. Computer disk and memory access efficiency may require special programming techniques to implement a full-sample data analysis. Advances in computing systems are alleviating these problems.

3.2 Implementation of Fay-BRR on the 1997-2005 NHIS

Lumley's *R* package *survey* **(**R Development Core Team, 2010) was used to compute replicate weights where each design year's quarterly poststratification weight-adjustment was an explicit part of the replication process. Due to the sizes of the accumulated 1997-2005 data files and the resulting data management and computational speed issues, we only applied the BRR construction to the NHIS sample adult databases (30K-40K sample adults per year) rather than to the NHIS sample person databases (90K-100K sample persons per year). The package computed a Hadamard matrix of dimension 348 for the 339 strata, thus 348 sets of replicate weights are produced. (The R *survey* package could not achieve the minimum required Hadamard dimension of 340.) For the adult sample there were over 1,500 weighting adjustment cells over the 9 years of data. Fay-factor adjustments in the range 0 to 0.90 by 0.10 units were used to provide replicate weights for our study, i.e., the paired Fay-multiplicative factors (f, 2-f) range from (0, 2.0) to (0.90, 1.10), with the former pair representing the traditional BRR weighting.

4. BRR Evaluations

4.1 Review of BRR Structural Results

The BRR estimated variance is computed as follows:
Let S be the index set covering all sample units in the NHIS,

$\{w^{(r)}_i\}_{i \varepsilon S}$ represents the survey weights for replicate set r, $r = 1,2,..,R$, and
$\{w^{(0)}_i\}_{i \varepsilon S}$ represents the survey weights for the original sample.

If θ is an estimator for a population parameter, Θ, using the weights $\{w^{(0)}_i\}_{i \varepsilon S}$, and $\theta(f)^{(r)}$ is an estimator for Θ using the replicate weights $\{w^{(r)}_i\}_{i \varepsilon S}$ defined using the Fay-factor, f, then the Fay-BRR estimated variance of θ is

$$var_{BRR}(\theta,f) = 1/(1-f)^2 \ \Sigma_{\{r=1 \text{ to } R\}} (\theta(f)^{(r)} - \theta)^2 / R.$$

The Rao and Shao (1999) paper presents several theoretical properties about the Fay-BRR estimated variance under four stated regularity conditions (Note, the ε in that paper is *(1-f)* in this paper). These conditions can be defined somewhat informally. It is assumed that we have a hypothetical population of first-stage clusters along with a sample of clusters both getting large, but with no sampling weight becoming disproportionately large. The *condition 2* implies that the characteristic to be estimated must be sufficiently "spread" over the clusters. The *condition 3* requires a 4^{th} moment stability. *Condition 4* requires a differentiability condition for a quantile. Smooth functions require a certain levels of continuous differentiability. Two interesting properties are

1. $var_{BRR}(\theta,f)/var(\theta) \rightarrow 1$, under asymptotic conditions for fixed f in [0, 1)
2. $var_{BRR}(\theta,f) \rightarrow var_{Linear}(\theta)$ as $f \rightarrow 1$

From ideal theoretical considerations, a Fay-BRR variance using a large f should behave like a linearized or jackknife variance estimator and favor smooth functions, while a small f should favor the non-smooth functions.

While the NHIS consists of a large sample, the choice of statistic, θ, the realized sampling imbalances, the choice of weighting adjustments, and many other realistic sampling and nonsampling design components, all result in deviations from ideal regularity conditions. In practice, the theoretical properties mentioned above may not be close to realization.

4.2 Fay-BRR Evaluations on the 1997-2005 NHIS

Our examination of the Fay-BRR variances for the NHIS was of limited scope for this study. The variable *Body Mass Index* (BMI), defined as weight/height^2 (kg/m^2), was considered. BMI should be skewed with a heavy right tail, and the extreme observations may noticeably influence the replicate estimates depending upon which PSUs are in the standard BBR replicate sample. Thus, BMI may be a good variable to test with the BRR options.

The behavior of the BRR variances depends in part upon the smoothness of the statistic along with the components of the sampling that may vary from application to application, e.g., domain estimates, and the dispersion of the sample within the strata and PSUs. For this work smooth estimators were exemplified by estimated means and regression parameters, and non-smooth estimators exemplified by sample percentiles.

In the examples that follow, some typical survey estimators are presented, and their standard errors are computed using the standard BRR and the Fay-BRR methods where the Fay-factor is taken at representative levels in the interval (0, 1). Comparisons of the different BRR methods on the NHIS are heuristic in nature. Since the true standard errors of targeted statistics are not known, it is difficult to assess any superiority of one form over another. The evaluations will focus on the totality of comparisons with an emphasis on patterns of behavior subject to sample sizes and dispersion of the data within strata and PSUs.

4.3 Examples:

4.3.1 Estimation of Mean BMI over a "Large" Domain

The estimated standard errors and year-to-year correlations covering NHIS years 1997-2005 for mean BMI for "non-black adults" was considered as a case where the sample sizes would be considered large, and the domain would cover most of the sampled PSUs. The regularity conditions discussed in Section 4.1 may be reasonably satisfied for this estimation situation as the selection of the Fay-factor had little impact on the magnitude of the Fay-BRR estimated standard errors. An examination of the Fay-BRR-estimated standard errors and year-to-year correlations for the Fay-factor, *f*, over the range 0.0 to 0.90 by 0.10 showed only negligible differences for these BMI-based estimates In Table 1 and Table 2, only the standard BRR computations with Fay-factor, *f*, =0 are displayed as the other Fay-factor results were quite similar. Similarly, for several other variables (not presented here), the standard errors for estimated means on data- and PSU-rich large domains changed very little as the Fay factor varied.

The estimated year-to-year correlations of Table 2 are a measure of the impact of the geographical clustering over time for the survey. In the past the sampled housing units had a greater proximity in adjacent years, and a decreasing magnitude of correlation might be expected over time. That pattern was broken during the 1997-2005 design to help reduce disclosure risk. This may help to explain the irregular pattern.

4.3.2 Estimation of Mean BMI over a "Small" Domain

The mean BMI for black males aged 18-25 years from 1997-2005 covers a small domain but includes many strata and PSUs. The results of Tables 3 and 4 correspond to Tables 1 and 2 as presented in the first example. As the Fay-factor varied on this smaller sample-size domain, changes were more noticeable. A gradual decline in the magnitude of the standard errors can be seen as the Fay-factor increased 0.0 to 0.90.

Some modest changes in estimated year-to-year correlation are also observed as the Fay-factor varied. For the Fay-factors 0.0 and 0.50 the estimated year-to-year correlation between 1998 and 2004 was 0.25 and 0.20, respectively. As there is no measure of the true variance, a conservative approach to selecting a Fay-factor for the estimated means would be to select a smaller Fay-factor.

The large positive and negative values in the year-to-year correlation of Table 4 may suggest instability of the quadratic forms used to produce the correlation matrix. It should be noted that estimating the "degrees of freedom" for a BRR variance estimator involves the estimation of a fourth moment. This paper does not address that issue.

4.3.3 BMI Regression over a "Small" Geographically Concentrated Domain

In this example, adults aged 65 years and older who resided in a domain consisting of 20 design strata were considered for a regression analysis using the 1997 NHIS. There were 438 observations in this domain. While estimated regression coefficients are smooth functions, this example should demonstrate BRR behavior under conditions far from the asymptotic conditions.

For each replicate weight in a Fay-factor generated replicate set, the model

```
BMI ~ 1 + age + age^2 + sex + race + current_smoker + sex×race,
```

with age transformed into a linear and quadratic orthogonal components, sex (male, female), race (black, nonblack), and current_smoker (yes, no), was fit using the R function *lm* with the fixed replicate weight. The estimated coefficients, the β's, and standard errors appear in Table 5. The linearization method's standard errors, produced with the R package *survey*, (treating the final adjusted weight as a sampling weight) are also displayed for comparison. In Table 6 the regression estimates and corresponding standard errors of expected BMI at age 70 for the different values of sex, race and smoking status, E(BMI | X), are provided.

The BRR estimates of the standard errors of the β's noticeably decrease, with some degree of stabilization as the Fay-factor increases. The linearization-based standard errors for intercept, age, race and current_smoker are smaller than the BRR-based versions, but for the sex and interaction sex×race standard errors the linearization-based versions are larger. For linearization we note that one typically assigns a *degrees of freedom* as the number of PSUs – number of strata, which in this example is 10. However, 7 regression parameters are also estimated. This would most likely reduce the degrees of freedom even more. (In the classical regression framework a degree of freedom is lost for each parameter estimated.) Linearized variance estimator instability may contribute to the lack of consistent behavior. If *t*-tests are computed for the β's, the significance of the magnitude of the *t*'s may be difficult to assess for several terms since "normal" asymptotic conditions are not realized. Only the current smoker with a $| t |$ always greater than 3.2 would seem to have evidence of significance.

The standard errors for the estimates of E (BMI | X) have decreasing behavior similar to that of the standard errors of the β's as the Fay-factor decreases. For the cases considered either standard error (linearization) ≤ standard error (BRR) or vice-versa.

In this "small sample" example recommendations cannot be easily made.

4.3.4 BMI Percentile Estimation over a Large Domain

For this example, the percentiles for BMI for adults in the 1998 NHIS (31,397 observations) at points 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95 and 99 along with corresponding standard errors were computed. This example should represent data- and PSU-rich situations. The BRR and linearization comparisons are shown in Table 7. This table shows large changes in the BRR standard errors as the Fay-factor, f, varies from 0.0 to 0.90. The BRR standard errors based on large Fay-factors, e.g., $f = 0.70$ and 0.90, behave very poorly, with some cases estimating the standard error as 0.0. This is consistent with jackknife behavior as discussed by Rao and Shao (1999). The position of the magnitude of the linearization-based standard error varies by percentile, thus no general statement on the relative magnitudes of Fay-BRR and the linearization can be made. As measured, the BMI variable has only 1866 distinct values to represent 31,397 truly continuous values. This high-degree of discretization may have had an impact on the interpolation of the replicate-weighted cumulative distribution function. For this work the R *survey* package was used for linearization, but Harell's R package *hmisc* was used for the BRR calculations. Furthermore, for linearization the final weights were treated as sampling weights, which also may have an impact on the estimates. Using the standard BRR method seems to provide the largest standard errors, and may be the most realistic in estimating the sampling variability for the non-smooth functions.

5. Conclusions

This paper represents a limited evaluation of the Fay-BRR method as applied to selected BMI-based statistics from the 1997-2005 NHIS adult samples. Additional work is still needed. Without some simulations from a universe and sampling methods similar to the NHIS, evaluation is somewhat speculative. Some limited conclusions follow.

As mentioned in Section 3.1 item 3, using a Fay factor in the computation of replicate poststratification adjustments reduces the possibility that an empty or small count adjustment cell will result in an undefined or highly inflated adjustment factor for the replicate weight. The full NHIS poststratification process tends to use "sufficiently" sampled poststratification cells. Thus, a re-poststratification based on traditional BRR replicate weights may not be adversely affected by being subjected to having many small post-cell sizes at replicate sample poststratification. The stability of this weighing feature may account for the negligible Fay-factor impact in example 1. For the NHIS Child sample (not considered here) small post-cell sizes may be present, and the use of a non-zero Fay-factor may reduce replicate-estimate variability.

For statistics based on "totals" which have a broad coverage over NHIS strata, the choice of the Fay-factor, *f*, seems to have a minor impact.

Percentile standard errors are greatly affected by choice of the Fay-factor, even for large samples. For percentiles, the traditional BRR may be the preferred one. More work is needed before recommending a single Fay-factor as a compromise.

Acknowledgements

The author thanks Kim Lochner for providing files for the NHIS poststratification cell classes and control totals that were used for the 1997-2005 years.

References

Botman, S., Moore, T., Moriarity, C., and Parsons, V. (2000). Design and Estimation for the National Health Interview Survey, 1995-2004, *Vital and Health Statistics,* 2(130).

Judkins, D. R. (1990). Fay's method for variance estimation. *Journal of Official Statistics*, 6 (3), 223-239.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org.

Rao, J.N.K. and Shao, J. (1999). Modified Balanced Repeated Replication for Complex Survey Data, *Biometrika*, 86, 403-415.

Wolter, K.M. (2007). *Introduction to Variance Estimation, Second Edition,* Springer

Table 1. Estimated Mean Body Mass Index (BMI) and Standard Errors[1] for Non-Black Adults from the 1997-2005 NHIS

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005
Mean	26.02	26.13	26.35	26.38	26.56	26.65	26.74	26.81	26.96
StdErr	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.04	0.05

[1]Standard BRR (Fay-factor = 0)

Table 2. Estimated Year-to-Year Correlations[1] of Estimated Mean Body Mass Index for Non-Black Adults from the 1997-2005 NHIS

Corr	1997	1998	1999	2000	2001	2002	2003	2004	2005
1997	1	.	.	.	.	.	.	.	.
1998	.22	1	.	.	.	.	.	.	.
1999	.08	.14	1	.	.	.	.	.	.
2000	.14	.05	.09	1	.	.	.	.	.
2001	.09	.10	.11	.15	1	.	.	.	.
2002	.14	.09	.02	.18	-.08	1	.	.	.
2003	.13	.20	.05	.11	.17	.14	1	.	.
2004	.17	.21	.06	.06	.02	.07	.20	1	.
2005	-.03	.11	.04	-.01	.01	.12	.12	.06	1

[1]Using Standard BRR (Fay-factor = 0)

Table 3. Estimated Mean BMI and Fay-BRR Standard Errors for Black males aged 18-25 years for the 1997-2005 NHIS

Means

97	98	99	00	01	02	03	04	05
26.21	26.34	25.46	25.49	26.14	26.42	25.55	26.82	25.78

Standard errors

f[1] \year	97	98	99	00	01	02	03	04	05
0.0	.39	.51	.56	.39	.47	.46	.43	.55	.41
0.1	.39	.51	.56	.39	.47	.45	.42	.54	.41
0.2	.39	.50	.55	.39	.46	.45	.42	.53	.40
0.3	.38	.50	.55	.38	.46	.45	.42	.53	.40
0.4	.38	.50	.54	.38	.46	.45	.41	.52	.40
0.5	.38	.50	.54	.38	.45	.44	.41	.52	.39
0.6	.38	.49	.54	.38	.45	.44	.41	.52	.39
0.9	.37	.49	.53	.37	.45	.44	.41	.51	.39

[1]Fay-factor

Table 4. Year-to-Year correlation 1997-2005 for Estimated Mean BMI for Black males aged 18-25 years for the NHIS

Standard BRR – with Fay factor 0

year	97	98	99	00	01	02	03	04	05
97	1	.	.	.	.	.	.	.	.
98	.13	1	.	.	.	.	.	.	.
99	.14	.16	1	.	.	.	.	.	.
00	-.02	-.12	.16	1	.	.	.	.	.
01	.17	.14	.14	-.12	1	.	.	.	.
02	-.15	.06	.09	-.06	.10	1	.	.	.
03	-.03	-.03	-.04	.02	-.02	.07	1	.	.
04	.09	.25[1]	-.01	-.03	-.07	-.01	-.02	1	.
05	.15	.05	-.22	.00	-.15	-.08	.09	.02	1

BRR-based with Fay factor 0.5

year	97	98	99	00	01	02	03	04	05
97	1	.	.	.	.	.	.	.	.
98	.11	1	.	.	.	.	.	.	.
99	.11	.17	1	.	.	.	.	.	.
00	-.02	-.16	.17	1	.	.	.	.	.
01	.15	.10	.11	-.11	1	.	.	.	.
02	-.17	.03	.00	-.07	.06	1	.	.	.
03	-.02	-.04	-.05	.03	.00	.05	1	.	.
04	.08	.20	-.06	-.05	-.05	.00	-.03	1	.
05	.16	-.04	-.19	.04	-.13	-.10	.14	.04	1

[1]Highlighted major differences

Table 5

Regression for the Model:

$BMI \sim 1 + age + age^2 + sex + race + cur.smoker + sex*race$

for 1997 NHIS adults aged 65+ years
restricted to 20 strata with 438 observations

β Estimates

	1	age[1]	age^2	sex	race	cur.smk	sex*race
	16.2	1436	-586	3.3	0.46	-2.22	-3.6

β Standard errors

	1	age[1]	age^2	sex	race	cur.smk	sex*race
Linear	9.3	1401	490	2.0	0.99	0.53	2.1
Fay							
0.0	10.4	1523	532	1.8	1.19	0.69	1.9
0.1	10.2	1502	525	1.8	1.14	0.68	1.8
0.3	10.0	1470	513	1.7	1.08	0.66	1.8
0.5	9.8	1446	505	1.7	1.05	0.65	1.7
0.7	9.7	1430	499	1.7	1.03	0.64	1.7
0.9	9.7	1420	496	1.6	1.02	0.63	1.7

[1]age was transformed to orthogonal quadratic form

Table 6

Regression for the Model:

BMI ~ 1 + age + age^2 + sex + race + cur.smoker + sex*race

for 1997 NHIS adults aged 65+ years
restricted to 20 strata with 438 observations

Estimated BMI for given covariates

E(BMI | X= [age=70 , sex, race, current.smoker])

	E(1)[1]	E(2)	E(3)	E(4)	E(5)	E(6)	E(7)	E(8)
E(BMI|X)	25.84	23.6	26.30	24.08	29.1	26.9	26.05	23.82

Standard errors

	SE(1)[1]	SE(2)	SE(3)	SE(4)	SE(5)	SE(6)	SE(7)	SE(8)
Linear	0.85	0.89	0.54	0.59	1.5	1.7	0.48	0.55
fay								
0.0	1.13	1.3	0.54	0.72	1.5	1.5	0.54	0.75
0.1	1.09	1.2	0.53	0.71	1.4	1.5	0.53	0.74
0.3	1.03	1.1	0.52	0.69	1.4	1.5	0.52	0.71
0.5	1.00	1.1	0.51	0.67	1.4	1.4	0.51	0.70
0.7	0.99	1.1	0.51	0.66	1.4	1.4	0.51	0.69
0.9	0.98	1.1	0.51	0.66	1.4	1.4	0.51	0.68

E( )[1]	Sex	race	current smoker	E( )	Sex	race	current smoker
1	M	Black	No	5	F	Black	No
2	M	Black	Yes	6	F	Black	Yes
3	M	nonBlack	No	7	F	nonBlack	No
4	M	nonBlack	Yes	8	F	nonBlack	Yes

Table 7 Percentiles and Standard Errors for BMI for adults in the 1998 NHIS

Percentile	1	5	10	20	30	40	50
BMI	17.64	19.49	20.58	22.13	23.42	24.39	25.7
				Standard Error			
Linear	0.10	0.046	0.026	0.051	0.036	0.010	0.031
f-BRR							
0.00	0.11	0.053	0.024	0.041	0.053	0.014	0.045
0.10	0.11	0.053	0.023	0.039	0.053	0.013	0.046
0.20	0.11	0.051	0.023	0.038	0.053	0.013	0.046
0.30	0.12	0.050	0.022	0.038	0.053	0.012	0.046
0.40	0.12	0.043	0.019	0.036	0.054	0.011	0.044
0.50	0.13	0.037	0.018	0.036	0.051	0.011	0.042
0.60	0.14	0.030	0.011	0.027	0.049	0.008	0.042
0.70	0.15	0.019	0.000	0.014	0.044	0.007	0.042
0.80	0.17	0.007	0.000	0.000	0.048	0.000	0.040
0.90	0.20	0.000	0.000	0.000	0.063	0.000	0.030

Percentile	60	70	80	90	95	99
BMI	26.63	28.27	30.12	33.21	36.24	43.83
			Standard Error			
Linear	0.036	0.031	0.054	0.12	0.14	0.28
f-BRR						
0.00	0.033	0.035	0.049	0.11	0.15	0.31
0.10	0.026	0.033	0.045	0.11	0.16	0.31
0.20	0.024	0.032	0.041	0.10	0.16	0.31
0.30	0.019	0.032	0.034	0.10	0.16	0.30
0.40	0.012	0.033	0.030	0.09	0.16	0.32
0.50	0.007	0.026	0.024	0.09	0.15	0.34
0.60	0.002	0.025	0.018	0.10	0.14	0.37
0.70	0.000	0.024	0.014	0.12	0.13	0.41
0.80	0.000	0.020	0.016	0.16	0.12	0.45
0.90	0.000	0.000	0.005	0.24	0.09	0.36