Weighting Survey Data: How To Identify Important Poststratification Variables

Michael P. Battaglia, Abt Associates Inc.; Martin R. Frankel, Abt Associates Inc. and Baruch College, CUNY; and Michael Link, Centers for Disease Control and Prevention

Abstract

Random-digit-dialing surveys such as the Behavioral Risk Factor Surveillance System (BRFSS) typically poststratify on age by gender cells for the adult population using control totals from an appropriate source such as the 2000 Census, the Current Population Survey (CPS), or private sector companies such as Claritas. Rao at al. (2005) used the 2000 Public Use Microdata Sample (PUMS) and the CPS to identify underrepresented sociodemographic subgroups in the BRFSS. This approach identifies potential poststratification variables on the basis of nonresponse. In our research we modeled key risk factor outcome variables rather than modeling nonresponse. Using logistic regression and CHAID we identified key "main effect" sociodemographic variables and important two-factor interactions. Using raking (Battaglia et al. 2005) we show how to incorporate several main effects and two-factor interactions into the weighting of the BRFSS survey data, and compare the resulting risk factor estimates with those based on the usual BRFSS weights.

Introduction

Survey researchers are increasingly concerned about potential bias in random-digit-dialing (RDD) surveys resulting from frame noncoverage and unit nonresponse. Households with no landline telephones, as well as those with only cellular telephones are excluded from the RDD sample frame (approximately 5 percent of the population). The ability of the population to move their telephone numbers almost anywhere in the country or to convert them into cellular telephones makes assessment of frame noncoverage at the sub-national level (e.g., state level) difficult because the RDD sample is drawn based on the area codes/central office codes. Unit nonresponse is an issue in any of the various survey modes (mail, telephone, in-person) but response rates to RDD surveys have been declining in the last decade (Curtin et al. 2005, Battaglia et al. 2006) in part due to growth in screening technologies, privacy concerns, telemarketing, and refusals.

In recognition of these issues, the Behavioral Risk factor Surveillance System has undertaken a program of research to evaluate alternative sampling frames for household surveys, examine multi-modality surveys, test the inclusion of cellular telephone adults in RDD samples, examine alternative weighting methodologies, and assess nonresponse bias in key risk factor estimates.

We first discuss previous research related to identifying factors related to nonresponse in a large RDD survey. We then report on our current research related to factors associated with key outcome variables in the same RDD survey. We show the results of incorporating additional factors in the weighting methodology for the survey and compare the results with the original weights developed for the survey.

Previous Research Examining Factors Related to Nonresponse

Rao at al. (2005) evaluated the degree to which noncoverage and unit nonresponse contribute to under-representation of important subgroups in RDD surveys. The Behavioral Risk Factor Surveillance System (BRFSS) -- a monthly RDD survey administered by all the states with assistance from the Centers for Disease Control and Prevention (CDC) to collect health-related information – was used as an example. BRFSS is an important survey, which generates state-specific prevalence estimates among adults (age 18+) of the major health conditions and behavioral risks associated with pre-mature morbidity and mortality. Details of the survey can be found in Mokdad et al. (2003) or at www.cdc.gov/brfss.

They were interested in evaluating noncoverage and nonresponse in six states (California, Illinois, North Carolina, New Jersey, Texas and Washington), which participated in a BRFSS pilot study designed to test techniques for improving coverage and reduce nonresponse (Link et al. 2005a, 2005b). Five of these states had experienced state-level response rates at or below 40% over the past several years (with North Carolina being the exception). From the 2003 BRFSS and the March 2003 Current Population Survey (CPS), Rao et al. identified the following socio-demographic variables of interest that were common to both surveys: age, sex, education, marital status, race/ethnicity, employment status, household income, number of children in household, type of household, and MSA versus non-MSA. Person weights were used to obtain the weighted frequencies. For the BRFSS, the person weight used did not include the final poststratification adjustments.

Rao et al. compared the distributions of the sociodemographic variables for six states from the 2003 BRFSS with the distribution of the same variables from the March 2003 CPS. They found that the youngest age group (*18-24*) was highly under-represented in NC, NJ, TX and WA. In CA and IL, they were under-represented but not by a substantial amount. Males were substantially under-represented in all six states. The least educated (*Did not graduate from high school*) were under-represented while the highly educated (*Graduated from college or technical school*) were over-represented. As would be expected, the magnitude of representation differed by state. Compared to the CPS, non-Hispanic whites were over-represented in all the states. Hispanics were under-represented in CA and TX, African-Americans were under-represented in IL, NC, NJ, and TX, and Asians were under-represented in all six states. Those who have never been married were under-represented in each of the six states while individuals who are married were over-represented in all states except CA. Those who are unemployed were over-represented in CA, NJ, TX and WA. The highest income category ($*50,000+*) was under-represented in all the states. In CA and TX the category *<$15,000* was over-represented while this income category was under-represented in all the other states. Compared to the CPS, there was an over-representation of households with no children. Households with only one woman were over-represented in all states except IL. Households with only 1 man and 1 woman were over-represented in CA and WA. Residence in an MSA was under-represented in CA and NJ, while it was over-represented in WA.

Identifying Factors Related to Key Survey Outcome Variables

Our current work relates to identifying sociodemographic factors associated with key risk factor dichotomous outcome variables in the 2003 BRFSS. We first identified 13 risk factor outcome variables to study (see Table 1).

Table 1: Thirteen BRFSS Risk Factor Outcome Variables (at risk versus not at risk)
HEALTH STATUS
HAVE HEALTH CARE COVERAGE
NO LEISURE TIME PHYSICAL ACTIVITY OR EXERCISE PAST
HIGH BLOOD PRESSURE RISK FACTOR
EVER TOLD BY DOCTOR YOU HAVE DIABETES
RISK FACTOR FOR RESPONDENTS AGED 65+ THAT HAD A FLU SHOT
CURRENT SMOKING STATUS RISK FACTOR.
HEAVY DRINKING RISK
1 Not At Risk
2 At Risk

BINGE DRINKING RISK FACTOR.
NO PHYSICAL ACTIVITY OR EXERCISE RISK FACTOR
EVER BEEN TESTED FOR HIV RISK FACTOR
RISK FACTOR FOR OVERWEIGHT OR OBESE
RISK FACTOR FOR LIFETIME ASTHMA PREVALENCE

An examination of the variables used by Rao et al. and a review of the sociodemographic variables available in the BRFSS resulted in our creating nine sociodemographic variables (see Table 2). We decided not to include household income in our analysis, because our ultimate objective was to add some additional sociodemographic variables to the BRFSS weighting methodology. As discussed below, this process involved using the CPS to create control totals for use in raking. Household income is generally subject to high item nonresponse rates, may be subject to considerable reporting error, and is typically measured very differently in a telephone survey asking a single income question versus determining income from all sources using several questions as is done in the CPS.

Table 2: Sociodemographic Variables in the 2003 BRFSS

AGEG5YR7	1 Age 18 to 24
	2 Age 25 to 34
	3 Age 35 to 44
	4 Age 45 to 54
	5 Age 55 to 64
	6 Age 65 to 74
	7 Age 75 plus
EDUCAG2	1 Did not graduate High School
	2 Graduated High School
	3 Attended College or Technical School
	4 Graduated from College or Technical School
EMPLOY_R	1 Unemployed
	2 Not Unemployed
CHLDCNT1	1 No children in household
	2 One child in household
	3 Two or more children in household
HTYPE1	1 HH with only 1 man
	2 HH with only 1 woman
	3 HH with only 1 man and 1 woman
	4 HH with more than 1 man and no women
	5 HH with more men than women
	6 HH with equal men and women
	7 HH with more than 1 woman and no men
	8 HH with more women than men

SEX	1 Male 2 Female
RACE2_R4	1 White only, Non-Hispanic 2 Black only, Non-Hispanic 3 Hispanic 4 All Others
MARITAL3	1 Married 2 Never married, member unmarried couple 3 Divorced, Widowed, Separated
MSANMSA	1 MSA residence 2 Non-MSA residence

Using the logistic regression procedure available in SAS, 13 weighted risk factor forward stepwise logistic regression models were run offering the 10 sociodemographic predictor variables. We focused on the key predictors in each model by identifying predictors that entered at the first, second or third step. Table 3 summarizes our findings. Age entered all three models in the first, second or third step. Education and race/ethnicity also entered most of the models. Marital status and gender entered 4 and 3 models, respectively.

Table 3: Key Predictor Variables in the 13 Logistic Regression Models

Variable	Number of Models
Age	13
Education	8
Race/ethnicity	9
Martial Status	4
Gender	3

In addition to these main effects we were also interested in identifying key two-factor interactions. This was accomplished with the 2003 BRFSS using weighted CHAID segmentation trees. We first collapsed some of the categories of the above five predictor variables: 1) age was collapsed into 3 categories (18-34, 35-54, and 55+), 2) education as collapsed into 2 categories (high school graduate or less, some college or more), and race/ethnicity was collapsed into 3 categories (nonHispanic white and other races, nonHispanic black, and Hispanic). Table 4 shows the key two-factor interactions that emerged from the CHAID analyses. Age by education was a key two-factor interaction in 4 of the 13 CHAID models.

Table 4: CHAID Results

Interaction	Number of CHAID Models
Age by education	4
Age by gender	3
Gender by race/ethnicity	2
Age by race/ethnicity	2
Education by marital status	2
Marital status by age	2
Marital status by gender	2
Education by race/ethnicity	1

Before proceeding to the discussion of adding variables to the BRFSS weighting methodology, we summarize our risk factor findings. We find that the risk factors are associated with age, education, race/ethnicity, marital status and gender. Rao et al. found that these variables are also related to nonresponse in the BRFSS. When this condition occurs there is potential for reducing nonresponse bias by incorporating such variables into the poststratification adjustments, specifically through the use of raking. In terms of two-factor interactions we decided to include age by education and age by race/ethnicity in the raking procedures described next.

Adding Variables to the BRFSS Weighting Methodology

The 2003 BRFSS weighting methodology involves the calculation of a base sampling weight (design weight) followed by poststratification to age 14 age (7 categories) by gender control totals or 28 age by gender by race/ethnicity (nonHispanic white versus all other race/ethnicity groups) to obtain the final weight. The control totals are obtained from Claritas. Our objective was to rake the 2003 BRFSS for each of the six states to CPS control totals constructed using the March 2002, 2003 and 2004 CPS. We combined three years of CPS data to add stability to the state-level control totals.

As one might expect the Claritas population distribution for age by gender or age by gender by race/ethnicity in a state did not agree exactly with the CPS distribution for 2003-2004. Before obtaining control totals from the CPS, we first took the CPS March supplement person weight for each year and divided it by three. We then ratio adjusted the CPS weight for the 14 age by gender or 28 age by gender by race/ethnicity categories, so that the CPS weighted counts were in agreement with the Claritas counts. This step was necessary because we wanted to compare the impact of adding additional variables to the BRFSS weighting with the results from using the

final BRFF weight. Once we had a new CPS weight, control totals were produced for race/ethnicity, education, marital status, age by education, and age by race/ethnicity. For each state we collapsed the race/ethnicity variable to combine small categories that had less than 5% of the BRFSS completed interviews in the state with another race/ethnicity category.

The CPS also has a variable indicating whether the household that the adult lives in has telephone service and so in each state we can estimate the number of adults living in nontelephone households at the time of the CPS interview. The 2003 BRFSS contains a variable indicating whether the respondent lives in a household that experiences an interruption in telephone service of a week or longer. Using the BRFSS design weight we estimated the percentage of adults in a state living in telephone households with an interruption in telephone service. Following the procedure described by Frankel et al. (2003) we then created a CPS control total margin for:

1. Adults in telephone households without an interruption in telephone service.
2. Adults in telephone households without an interruption in telephone service and adults living in nontelephone households.

The inclusion of the nontelephone margin in the raking is intended to compensate for noncoverage from the exclusion of adults living in nontelephone households.

For each of the 13 risk factor outcome variables, we used the BRFSS design weight and the BRFSS final weight to estimate the percent of adults with a risk factor in each of the six states. We then used a SAS raking macro (Battaglia et al. 2005) to create 10 new weights for the BRFSS in each of the six states. The details of the margins included in each raking are shown in Table 5. The logic to the ordering of the 10 rakings is as follows: 1) the first 5 raking do not include a nontelephone adjustment using the interruption margin described above, 2) most survey statisticians would give highest priority to include a detailed race/ethnicity margin, even if a state has an age by gender by race/ethnicity margin that limited to nonHispanic white versus all other race/ethnic groups, 3) based on the logistic regression modeling results education will next be entered as a margin, followed by marital status, and 4) based on the CHAID results the age by education two-variable margin will next be entered and finally the age by race/ethnicity two-variable margin will be entered into the raking. Keep in mind that our findings are sensitive to the order of entry of the margins.

Table 5: 10 BRFSS Rakings

Without interruption in telephone service margin:	
1. Age by gender or age by gender by race/ethnicity	And race/ethnicity
2. Age by gender or age by gender by race/ethnicity and race/ethnicity	And education
3. Age by gender or age by gender by race/ethnicity, race/ethnicity, education	And marital status
4. Age by gender or age by gender by race/ethnicity, race/ethnicity and marital status	And age by education
5. Age by gender or age by gender by race/ethnicity, race/ethnicity and age by education	And age by race/ethnicity
With interruption in telephone service margin:	
6. Age by gender or age by gender by race/ethnicity	And race/ethnicity and interruption in telephone service
7. Age by gender or age by gender by race/ethnicity and race/ethnicity	And education and interruption in telephone service
8. Age by gender or age by gender by race/ethnicity, race/ethnicity, education	And marital status and interruption in telephone service
9. Age by gender or age by gender by race/ethnicity, race/ethnicity and marital status	And age by education and interruption in telephone service
10. Age by gender or age by gender by race/ethnicity, race/ethnicity and age by education	And age by race/ethnicity and interruption in telephone service

All of the rakings converged quickly (less than 10 iterations) using a convergence criterion of 1.0.

Results

We show the results of the 10 rakings for two states – California and Texas. California uses age by gender by race/ethnicity poststratification, and based on the CPS has only 2.8% of adults residing in nontelephone households. The Texas BRFSS used age by gender poststratification, and based on the CPS has a higher percent of adults, 5.7%, residing in nontelephone households. The race/ethnicity margin that we created using the 5% rule for Texas contains three categories – nonHispanic white, nonHispanic black, and Hispanic plus nonHispanic other races. For California the race/ethnicity margin contains four categories -- nonHispanic white, nonHispanic black, Hispanic, and nonHispanic other races.

Tables 6 and 7 show the resulting risk factor estimates and standard errors obtained from SUDAAN. We will concentrate on three key risk factors – general health, health insurance

status, and current smoking status. In Figures 1 to 6, we show the estimates for California and Texas.

In California the addition of the race/ethnicity margin has a small effect of the three risk factor estimates. The raking that includes race/ethnicity and adds education sharply raises all three risk factor estimates. In addition of marital status, age by education, and age by race/ethnicity causes little further change in the estimates. Furthermore, the inclusion on the nontelephone margin in the raking has little impact on the three risk factor estimates (no impact at all on the current smoking estimates). Compared to the risk factor estimates based on the final weight, the risk factor estimates from raking #10, which includes the nontelephone margin and the age by race margin, the three estimates increase by 9.9%, 6.2%, and 6.0%, respectively.

In Texas the addition of the race/ethnicity margin has a larger effect of the three risk factor estimates. The raking that includes race/ethnicity and adds education sharply raises all three risk factor estimates. In addition of marital status, age by education, and age by race/ethnicity causes a small additional change in the estimates. Furthermore, the inclusion on the nontelephone margin in the raking noticeably raises all the three risk factor. Compared to the risk factor estimates based on the final weight, the risk factor estimates from raking #10, which includes the nontelephone margin and the age by race margin, the three estimates increase by 14.9%, 10.9%, and 4.1%, respectively. In general, we find that the inclusion of additional variables in the raking raises the risk factor estimates, in other words, weighting on age by gender or age by gender by a two-category race/ethnicity variables tends to under-estimate risk factor levels.

Table 6: California Raking Results

	BRFSS Final Weight	SE	Add Race/ethnicity	SE	Add Education	SE	Add Marital status	SE	Add Age by Educ	SE	Add Age by Race/ethnicity	SE
Health Status												
Without Nontelephone Adjustment	15.1	0.675	14.8	0.669	16.3	0.753	16.4	0.757	16.5	0.758	16.5	0.758
With Telephone Adjustment	15.1	0.675	15.0	0.685	16.5	0.763	16.6	0.766	16.6	0.767	16.6	0.767
Health Care Coverage												
Without Nontelephone Adjustment	16.1	0.733	15.7	0.726	16.8	0.791	16.9	0.798	16.8	0.795	16.8	0.795
With Telephone Adjustment	16.1	0.733	16.0	0.749	17.1	0.814	17.2	0.820	17.1	0.816	17.1	0.816
No Leisure Time Activity or Exercise												
Without Nontelephone Adjustment	22.3	0.798	22.1	0.798	23.5	0.863	23.5	0.864	23.5	0.864	23.5	0.864
With Telephone Adjustment	22.3	0.798	22.4	0.818	23.8	0.883	23.8	0.884	23.8	0.883	23.8	0.883
High Blood Pressure												
Without Nontelephone Adjustment	23.4	0.747	23.4	0.747	23.9	0.792	23.9	0.792	24.0	0.796	24.0	0.796
With Telephone Adjustment	23.4	0.747	23.4	0.756	23.9	0.794	23.9	0.795	23.9	0.799	23.9	0.799
Ever Told By Doctor You Have Diabetes												
Without Nontelephone Adjustment	13.4	2.006	13.1	1.951	13.5	2.045	13.3	2.028	13.2	2.014	13.2	2.014
With Telephone Adjustment	13.4	2.006	12.9	1.911	13.3	1.998	13.1	1.985	13.0	1.969	13.0	1.969

	BRFSS Final Weight	SE	Add Race/ethnicity	SE	Add Education	SE	Add Marital status	SE	Add Age by Educ	SE	Add Age by Race/ethnicity	SE
Respondents 65+ Flu Shot												
Without Nontelephone Adjustment	27.5	1.931	27.7	1.959	27.7	2.039	27.7	2.034	27.6	2.048	27.6	2.048
With Telephone Adjustment	27.5	1.931	27.9	1.971	27.8	2.045	27.8	2.040	27.7	2.055	27.7	2.055
Current Smoking Status												
Without Nontelephone Adjustment	16.8	0.698	16.9	0.701	17.8	0.752	17.9	0.754	17.8	0.752	17.8	0.752
With Telephone Adjustment	16.8	0.698	16.9	0.706	17.8	0.760	17.9	0.761	17.8	0.759	17.8	0.759
Heavy Drinking												
Without Nontelephone Adjustment	5.7	0.409	5.6	0.400	5.7	0.436	5.8	0.438	5.7	0.435	5.7	0.435
With Telephone Adjustment	5.7	0.409	5.6	0.396	5.7	0.428	5.7	0.430	5.7	0.428	5.7	0.428
Binge Drinking												
Without Nontelephone Adjustment	15.9	0.701	15.7	0.692	15.7	0.718	15.8	0.720	15.8	0.722	15.8	0.722
With Telephone Adjustment	15.9	0.701	15.6	0.690	15.6	0.715	15.6	0.716	15.7	0.718	15.7	0.718
No Physical Activity or Exercise												
Without Nontelephone Adjustment	7.9	0.548	7.9	0.555	8.5	0.620	8.5	0.623	8.5	0.621	8.5	0.621
With Telephone Adjustment	7.9	0.548	7.9	0.561	8.5	0.624	8.6	0.626	8.5	0.624	8.5	0.624
Ever Been Tested for HIV												
Without Nontelephone Adjustment	50.6	1.026	50.7	1.031	50.9	1.066	51.0	1.067	51.0	1.066	51.0	1.066
With Telephone Adjustment	50.6	1.026	50.7	1.042	50.9	1.074	50.9	1.075	51.0	1.074	51.0	1.074

	BRFSS Final Weight	SE	Add Race/ethnicity	SE	Add Education	SE	Add Marital status	SE	Add Age by Educ	SE	Add Age by Race/ethnicity	SE
Overweight or Obese												
Without Nontelephone Adjustment	59.3	0.900	58.6	0.913	59.4	0.941	59.3	0.943	59.3	0.943	59.3	0.943
With Telephone Adjustment	59.3	0.900	58.5	0.922	59.2	0.949	59.2	0.950	59.2	0.951	59.2	0.951
Lifetime Asthma Prevalence												
Without Nontelephone Adjustment	13.4	0.606	13.5	0.615	13.6	0.639	13.6	0.641	13.7	0.643	13.7	0.643
With Telephone Adjustment	13.4	0.606	13.6	0.626	13.7	0.647	13.7	0.648	13.7	0.650	13.7	0.650

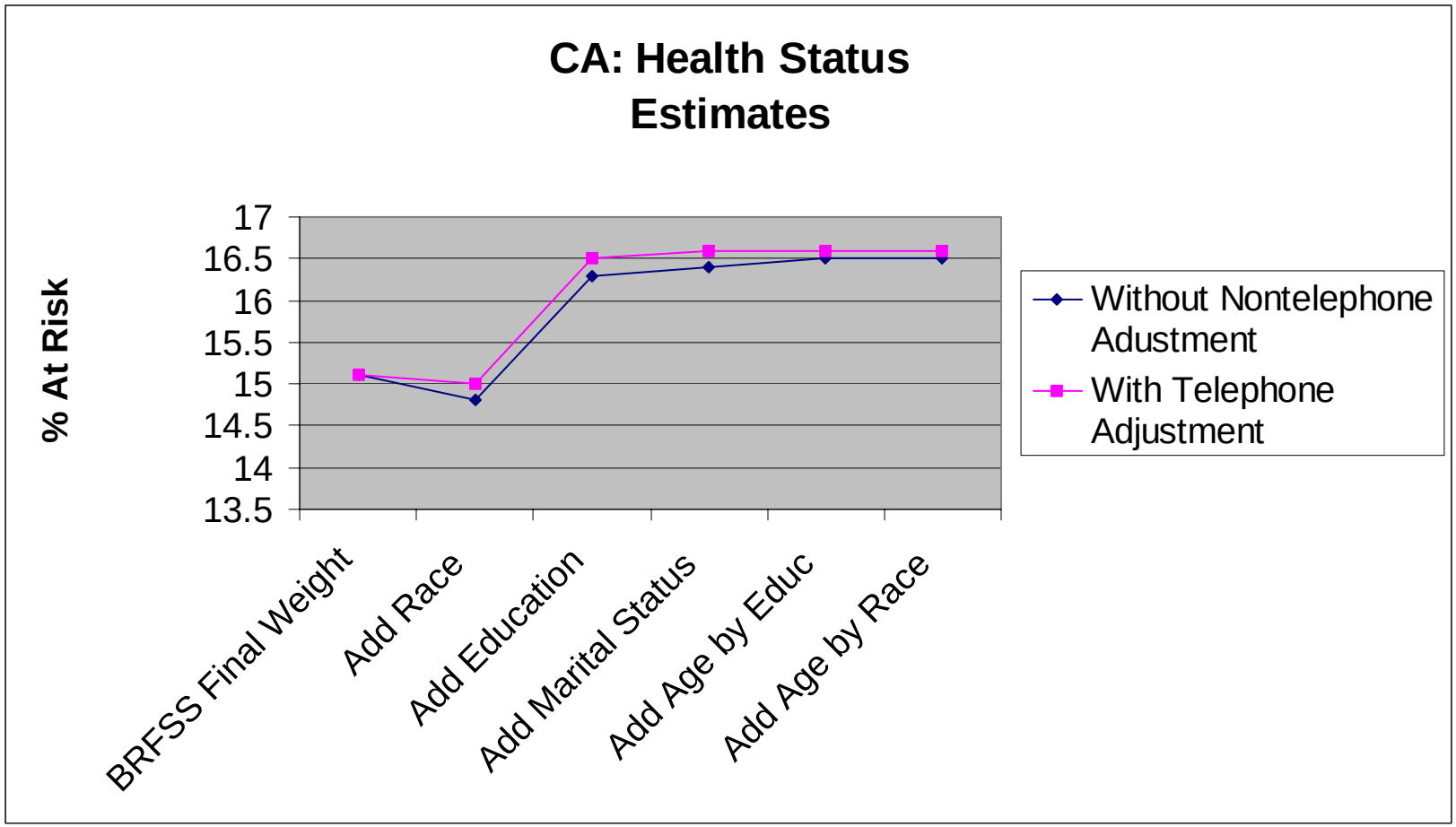
Table 7: Texas Raking Results

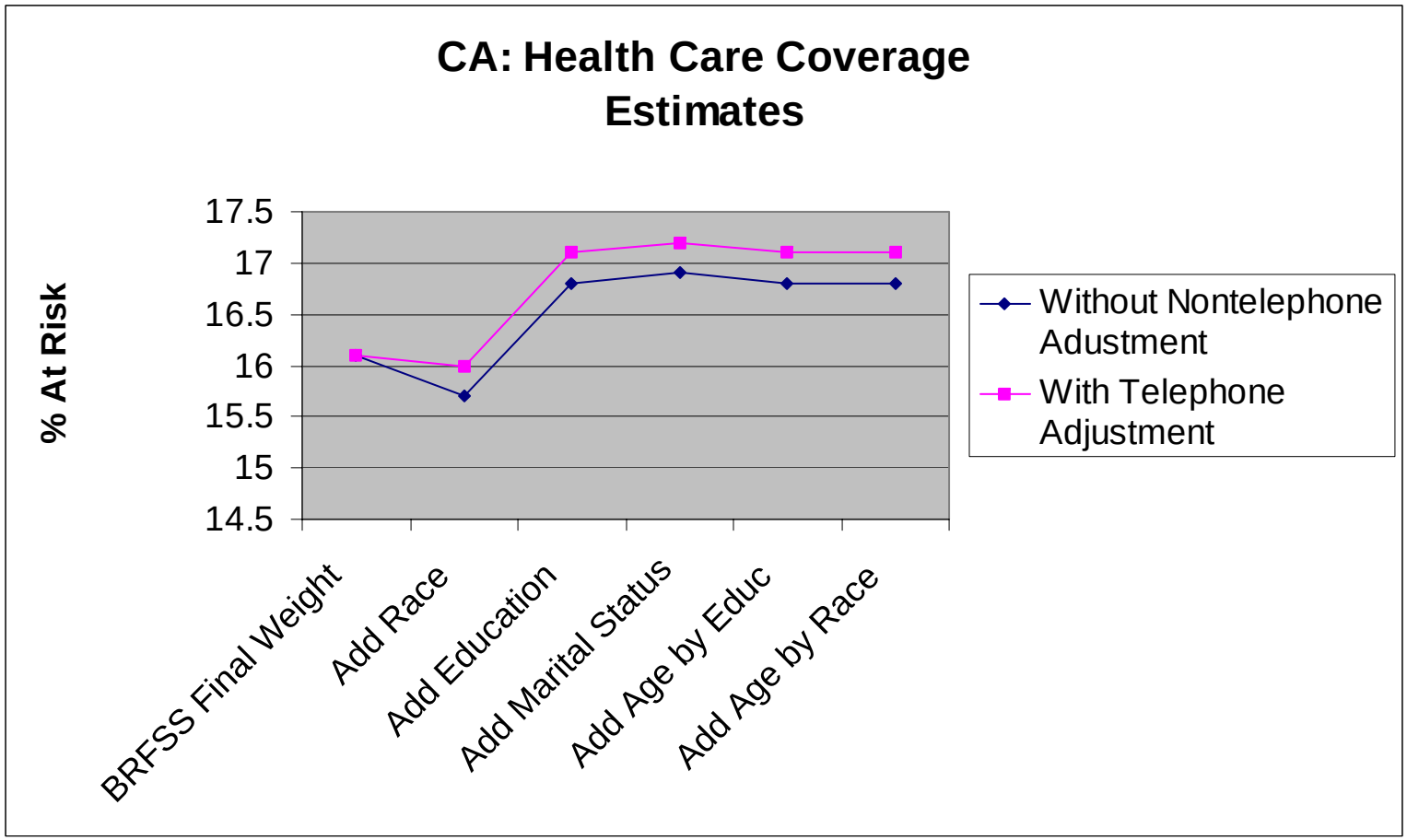
	BRFSS Final Weight	SE	Add Race/ethnicity	SE	Add Education	SE	Add Marital status	SE	Add Age by Educ	SE	Add Age by Race/ethnicity	SE
Health Status												
Without Nontelephone Adjustment	20.2	0.609	21.1	0.645	22.2	0.682	22.3	0.681	22.4	0.681	22.4	0.683
With Telephone Adjustment	20.2	0.683	22.1	0.698	23.0	0.728	23.1	0.727	23.1	0.726	23.2	0.727
Health Care Coverage												
Without Nontelephone Adjustment	26.7	0.716	28.0	0.751	29.3	0.776	29.4	0.778	29.2	0.774	29.1	0.772
With Telephone Adjustment	26.7	0.772	28.7	0.784	29.7	0.804	29.9	0.804	29.6	0.801	29.6	0.799
No Leisure Time Activity or												

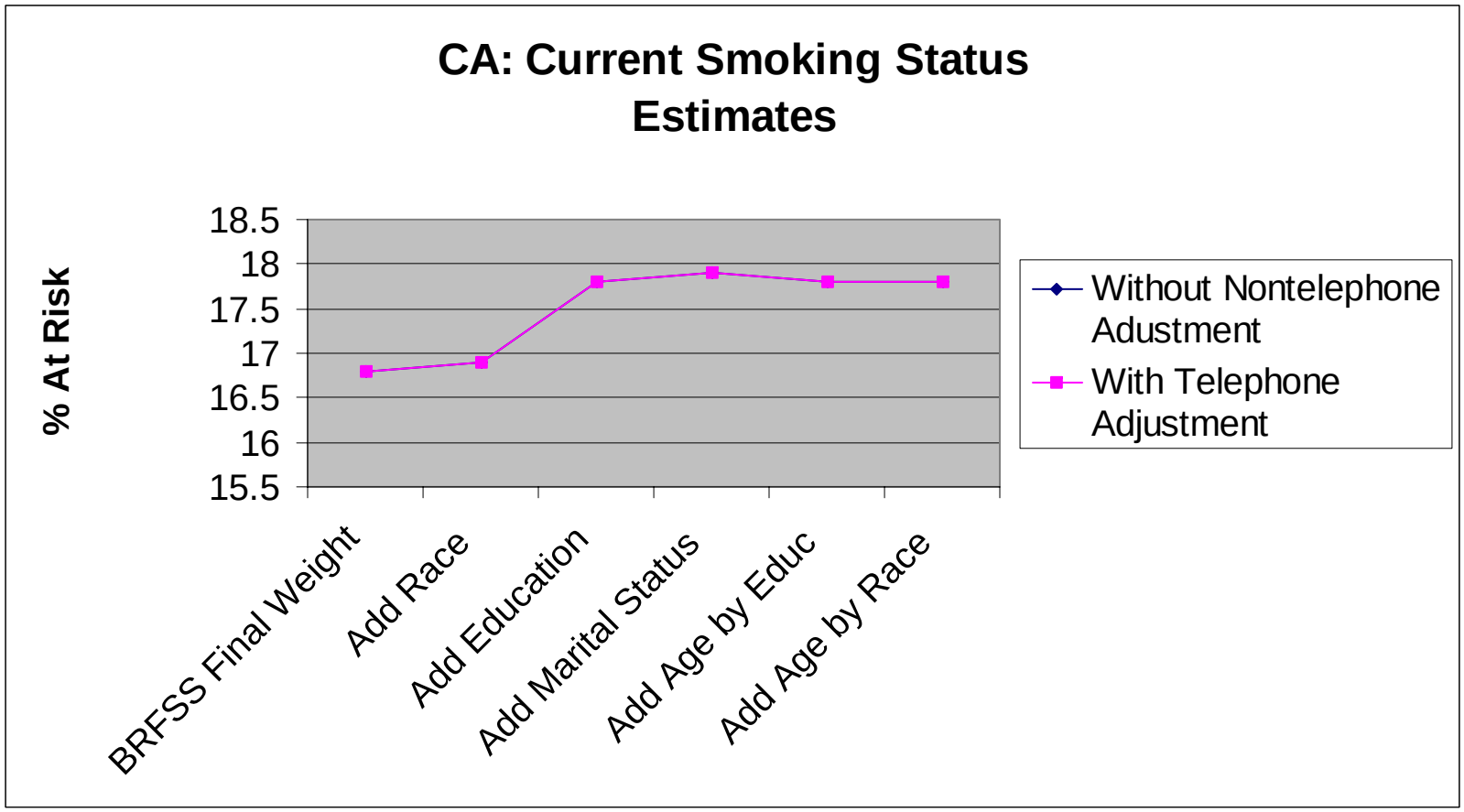
	BRFSS Final Weight	SE	Add Race/ethnicity	SE	Add Education	SE	Add Marital status	SE	Add Age by Educ	SE	Add Age by Race/ethnicity	SE
Exercise												
Without Nontelephone Adjustment	27.6	0.683	28.6	0.716	29.8	0.743	29.8	0.742	29.8	0.742	29.7	0.741
With Telephone Adjustment	27.6	0.741	29.1	0.744	30.0	0.766	30.1	0.765	30.1	0.765	30.0	0.764
High Blood Pressure												
Without Nontelephone Adjustment	24.6	0.620	24.9	0.644	25.2	0.660	25.2	0.660	25.2	0.663	25.2	0.664
With Telephone Adjustment	24.6	0.664	25.2	0.668	25.4	0.683	25.5	0.683	25.5	0.687	25.4	0.687
Ever Told By Doctor You Have Diabetes												
Without Nontelephone Adjustment	11.2	1.386	11.0	1.386	10.8	1.380	10.7	1.368	10.5	1.346	10.3	1.319
With Telephone Adjustment	11.2	1.319	10.5	1.355	10.4	1.352	10.3	1.352	10.1	1.323	9.9	1.295
Respondents 65+ Flu Shot												
Without Nontelephone Adjustment	32.3	1.546	32.7	1.585	33.1	1.623	33.2	1.616	33.8	1.678	34.1	1.730
With Telephone Adjustment	32.3	1.730	32.9	1.604	33.2	1.633	33.2	1.628	33.9	1.697	34.2	1.749
Current Smoking Status												
Without Nontelephone Adjustment	22.1	0.656	21.7	0.660	22.4	0.683	22.6	0.687	22.4	0.681	22.5	0.680
With Telephone Adjustment	22.1	0.680	22.3	0.693	23.0	0.714	23.1	0.716	22.9	0.710	23.0	0.710
Heavy Drinking												
Without Nontelephone Adjustment	5.9	0.393	5.8	0.401	5.8	0.402	5.9	0.407	5.9	0.408	5.9	0.408
With Telephone Adjustment	5.9	0.408	5.9	0.401	5.9	0.403	5.9	0.409	5.9	0.409	5.9	0.410

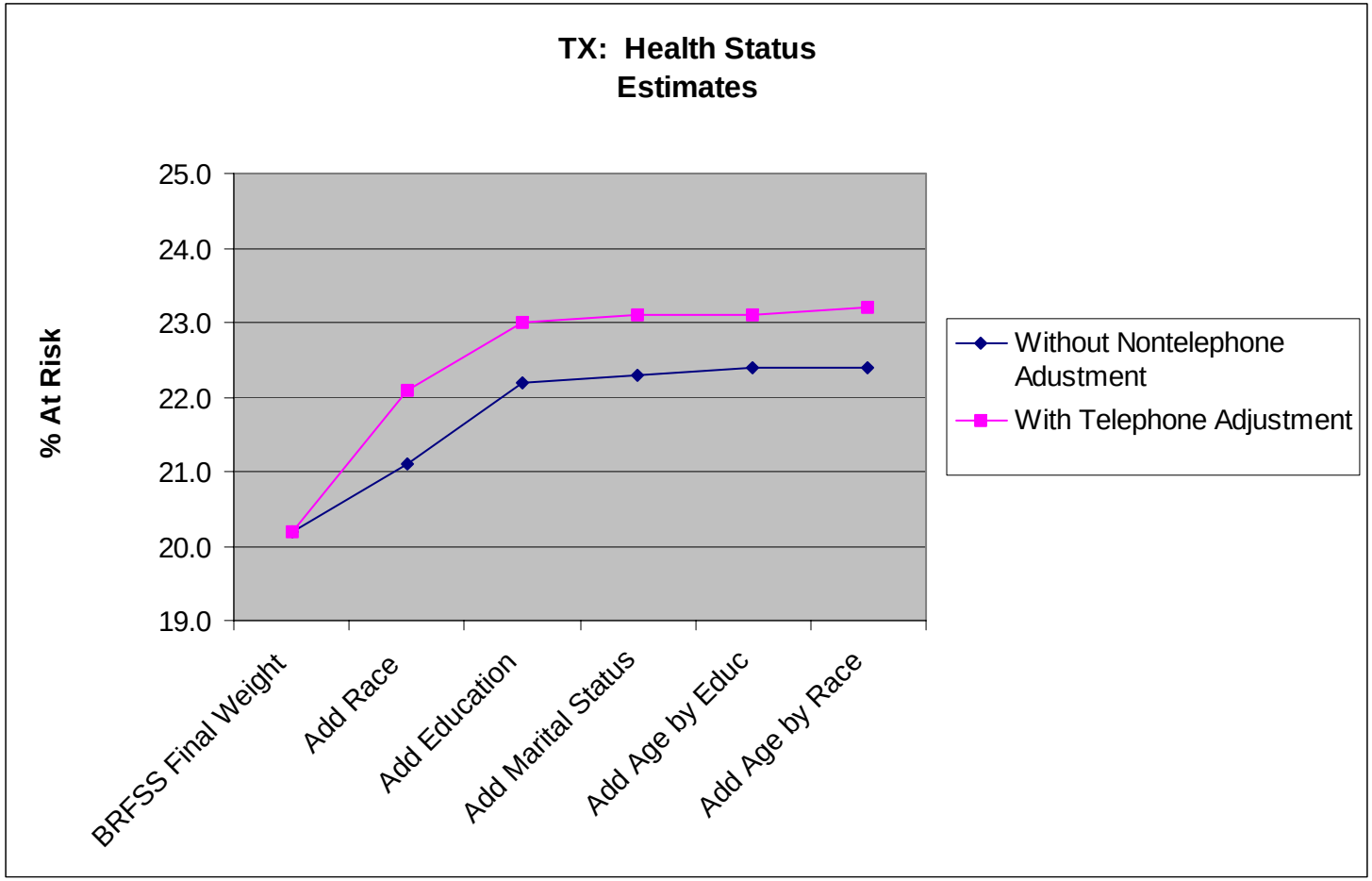
	BRFSS Final Weight	SE	Add Race/ethnicity	SE	Add Education	SE	Add Marital status	SE	Add Age by Educ	SE	Add Age by Race/ethnicity	SE
Binge Drinking												
Without Nontelephone Adjustment	16.3	0.610	16.2	0.621	16.3	0.631	16.4	0.634	16.4	0.632	16.4	0.632
With Telephone Adjustment	16.3	0.632	16.5	0.643	16.6	0.651	16.7	0.655	16.7	0.654	16.7	0.654
No Physical Activity or Exercise												
Without Nontelephone Adjustment	11.1	0.496	11.9	0.539	12.5	0.572	12.5	0.571	12.5	0.570	12.5	0.569
With Telephone Adjustment	11.1	0.569	12.1	0.557	12.6	0.583	12.6	0.583	12.6	0.582	12.6	0.582
Ever Been Tested for HIV												
Without Nontelephone Adjustment	52.6	0.873	52.3	0.898	52.6	0.912	52.6	0.913	52.6	0.910	52.6	0.909
With Telephone Adjustment	52.6	0.909	51.9	0.921	52.2	0.934	52.2	0.934	52.2	0.933	52.2	0.931
Overweight or Obese												
Without Nontelephone Adjustment	61.5	0.770	62.5	0.781	62.9	0.790	62.8	0.790	62.9	0.791	62.7	0.794
With Telephone Adjustment	61.5	0.794	62.8	0.792	63.1	0.800	63.0	0.800	63.1	0.801	63.0	0.804
Lifetime Asthma Prevalence												
Without Nontelephone Adjustment	11.3	0.491	11.1	0.505	11.0	0.508	11.1	0.511	11.1	0.510	11.1	0.512
With Telephone Adjustment	11.3	0.512	11.2	0.506	11.1	0.509	11.1	0.513	11.1	0.512	11.2	0.515

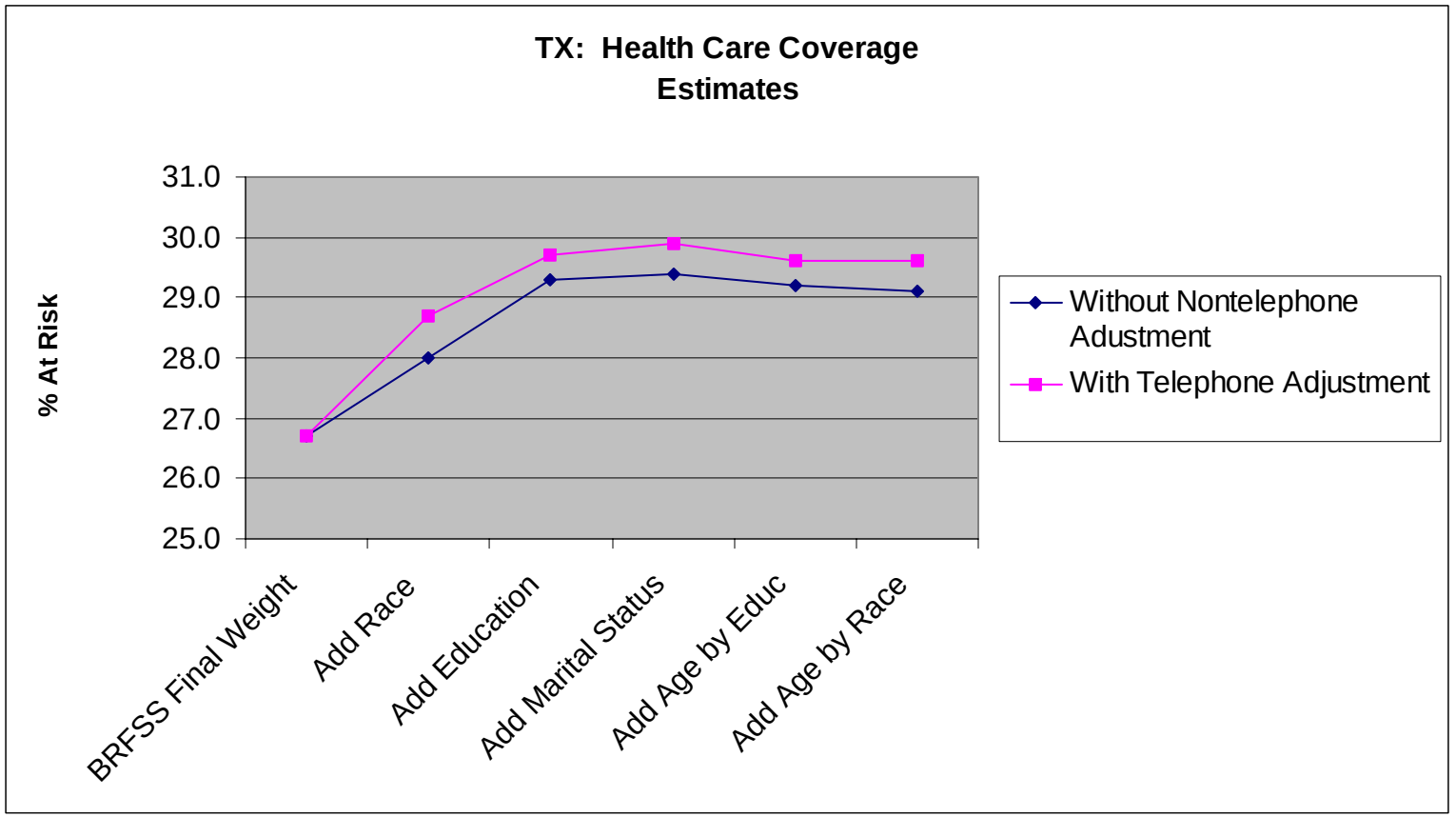
Figures 1-6:

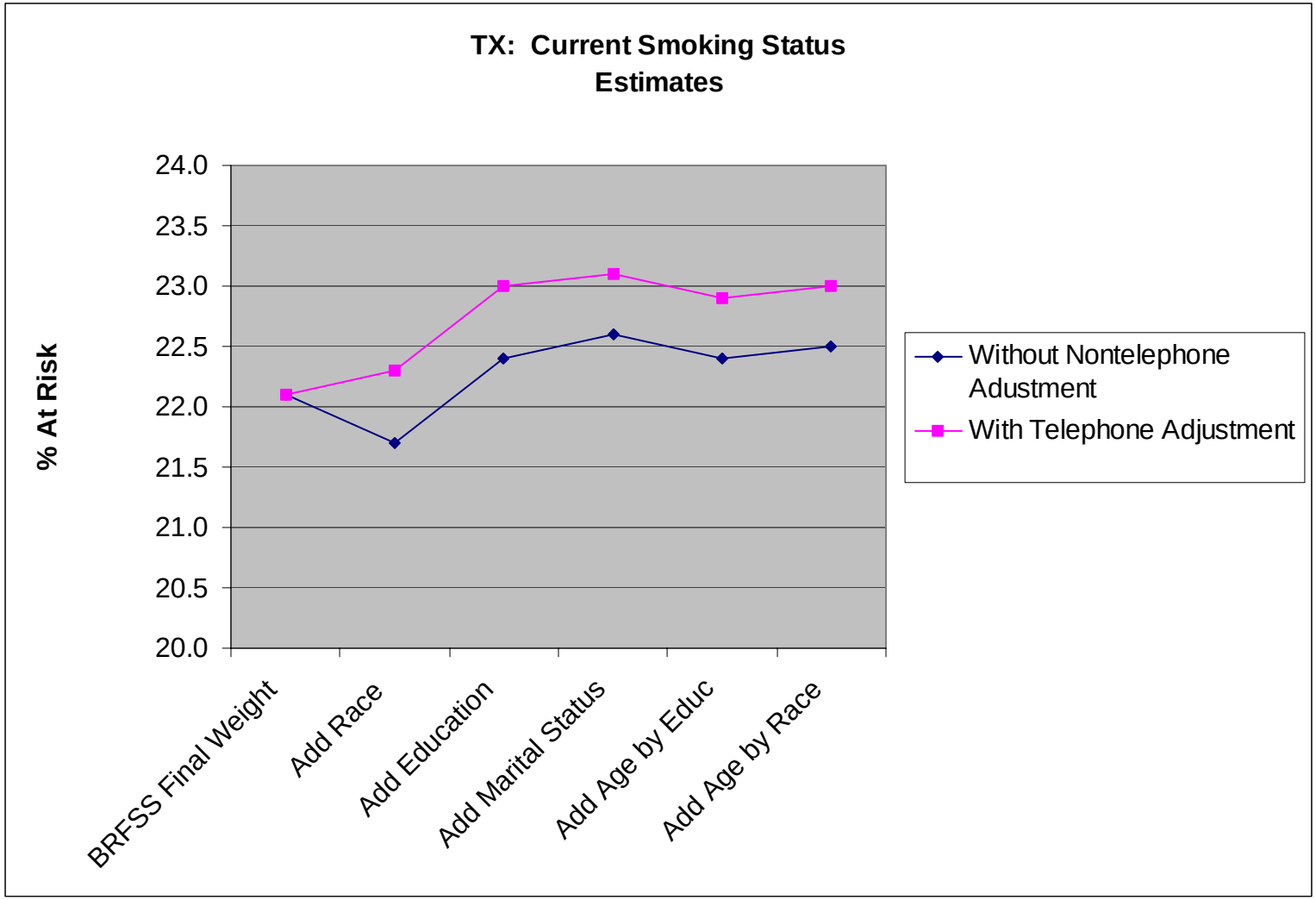

CA: Health Status Estimates
% At Risk
17
16.5
16
15.5
15
14.5
14
13.5
BRFSS Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Without Nontelephone Adustment
With Telephone Adjustment


CA: Health Care Coverage Estimates
% At Risk
17.5
17
16.5
16
15.5
15
14.5
BRFSS Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Without Nontelephone Adustment
With Telephone Adjustment


CA: Current Smoking Status Estimates
% At Risk
18.5
18
17.5
17
16.5
16
15.5
BRFSS Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Without Nontelephone Adustment
With Telephone Adjustment

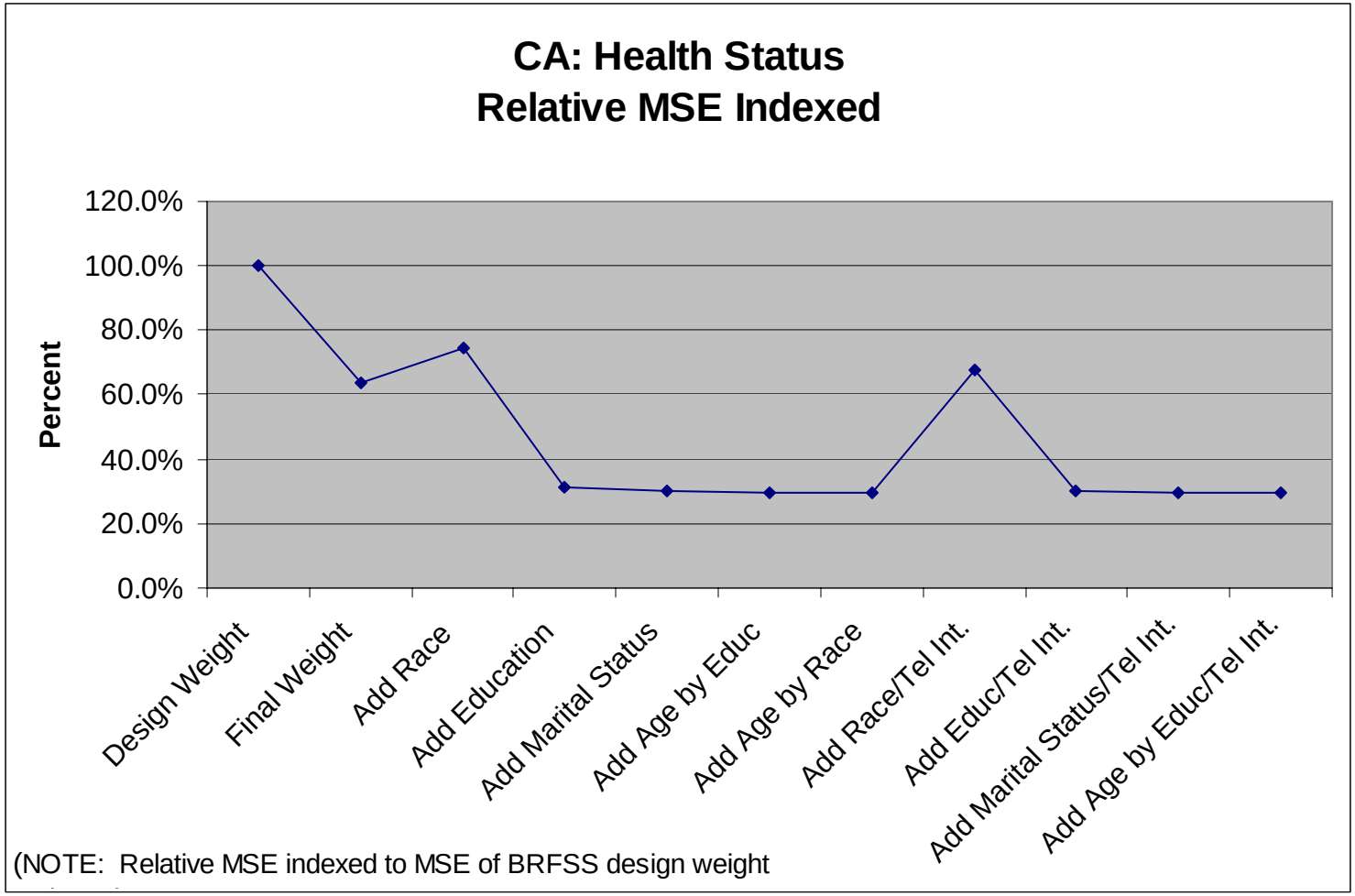

TX: Health Status Estimates
% At Risk
25.0
24.0
23.0
22.0
21.0
20.0
19.0
BRFSS Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Without Nontelephone Adustment
With Telephone Adjustment

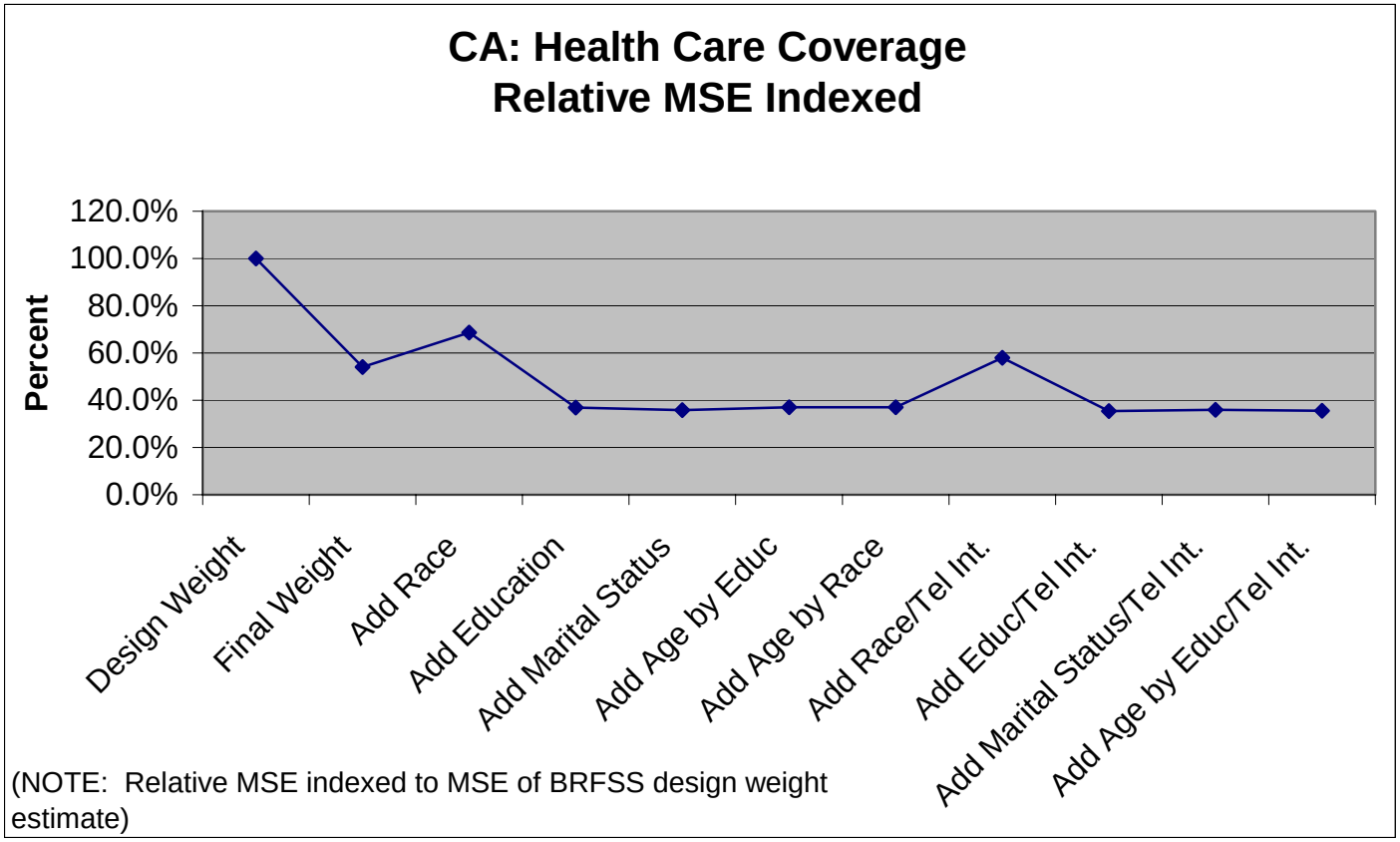

TX: Health Care Coverage Estimates
% At Risk
31.0
30.0
29.0
28.0
27.0
26.0
25.0
BRFSS Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Without Nontelephone Adustment
With Telephone Adjustment


TX: Current Smoking Status Estimates
% At Risk
24.0
23.5
23.0
22.5
22.0
21.5
21.0
20.5
20.0
BRFSS Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Without Nontelephone Adustment
With Telephone Adjustment

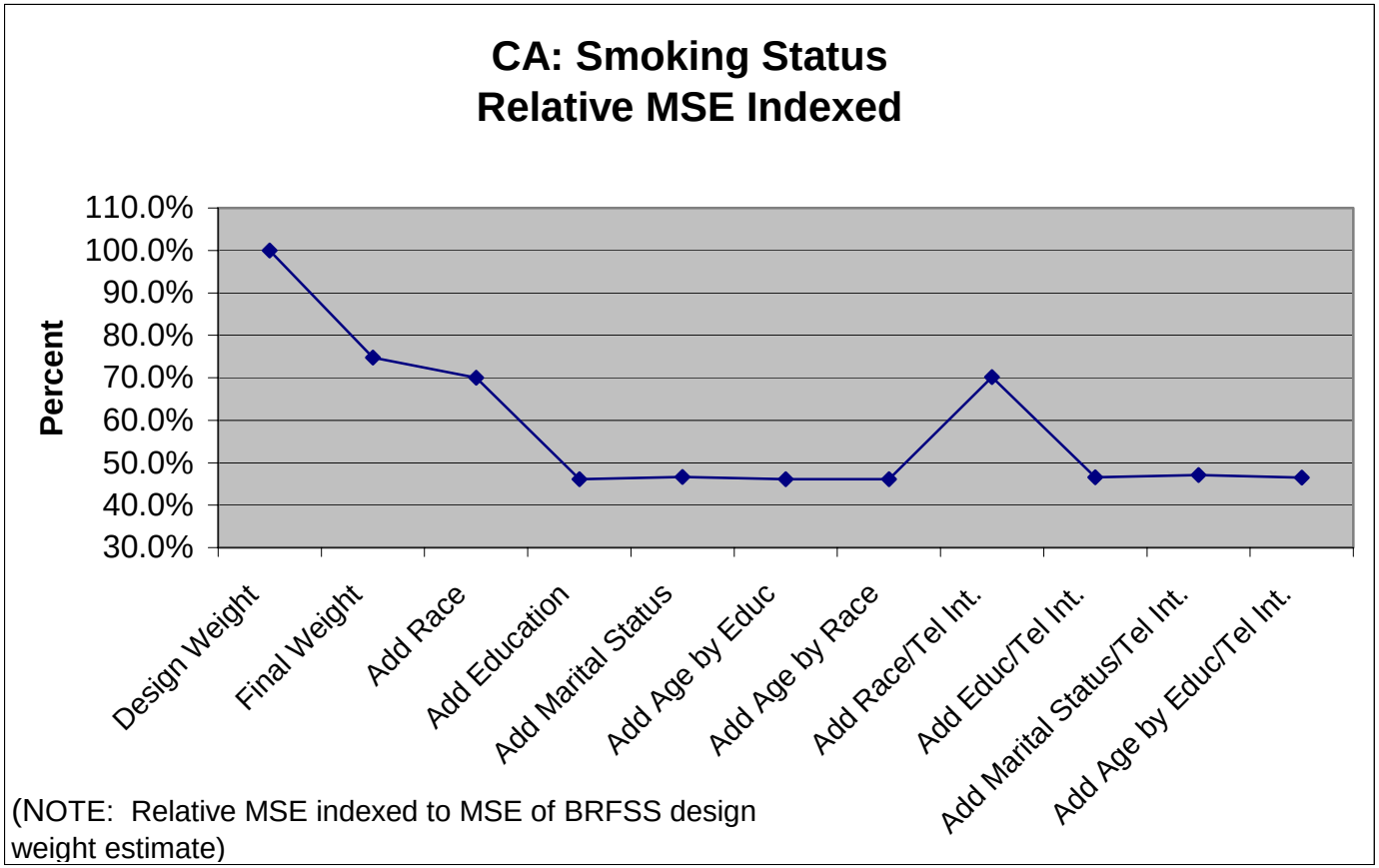
We developed estimates of the mean squared error of the risk factor estimates (based on the design weight, the final weight, and raking weights #1 to #9) by treating the estimates from raking #10 as unbiased. Relative mean squared error estimates were calculated by dividing the square root of the mean squared error estimates by the risk factor estimate from raking #10. Finally, we indexed the relative mean squared error estimates to the relative mean squared error estimates resulting from the design weight. The indexed relative MSE results are shown in Figures 6-12. By definition the indexed relative MSE for the design weight estimates is 100%. Because the inclusion of more variables in the raking typically increases the variance, it is possible for the indexed relative MSE for estimates based on one of the other weights to exceed 100%. For California the estimates based on the final weight and those for raking #1 (includes race/ethnicity) yield a reduction in the indexed relative MSE. However, a large additional reduction is seen with the addition of education to the raking. The inclusion of the nontelephone adjustment margin in the raking has very little impact on the indexed relative MSE in California. We see a similar pattern in Texas except in terms of the indexed relative MSE for the final weight and the raking that includes race/ethnicity. Similar to California we see that the addition of education to the raking causes a large drop in the indexed relative MSE. However, unlike California, the inclusion of the nontelephone adjustment margin has a noticeable impact on further reducing the indexed relative MSE. For general health status and health insurance status, the value of the indexed relative MSE is 30% or lower for the raking that includes the nontelephone margin and the age by education margin (raking #9). The inclusion of education, a socioeconomic status variable, is clearly important, however, the inclusion of the nontelephone adjustment margin in the raking can also be important.

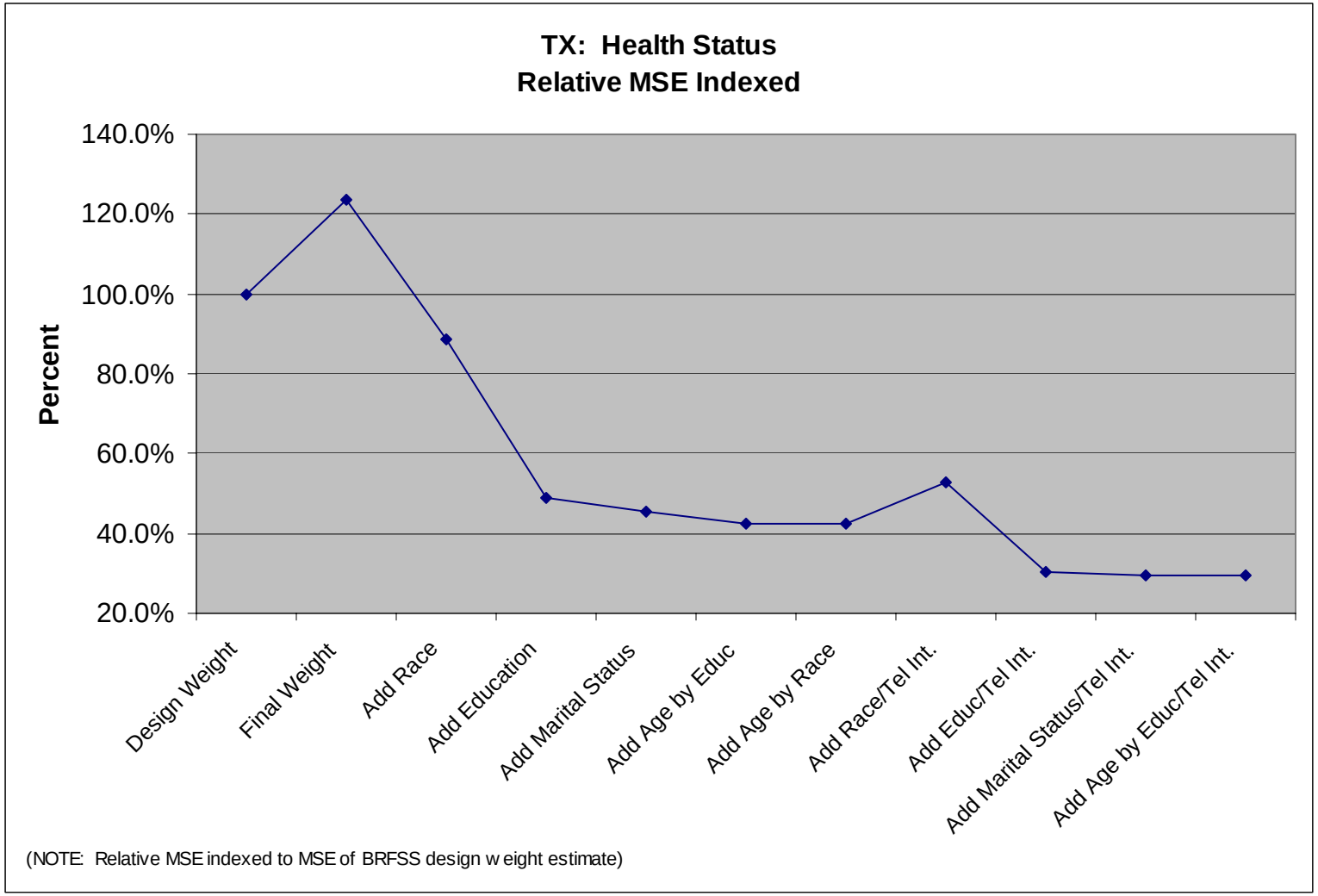
Figures 6-12:

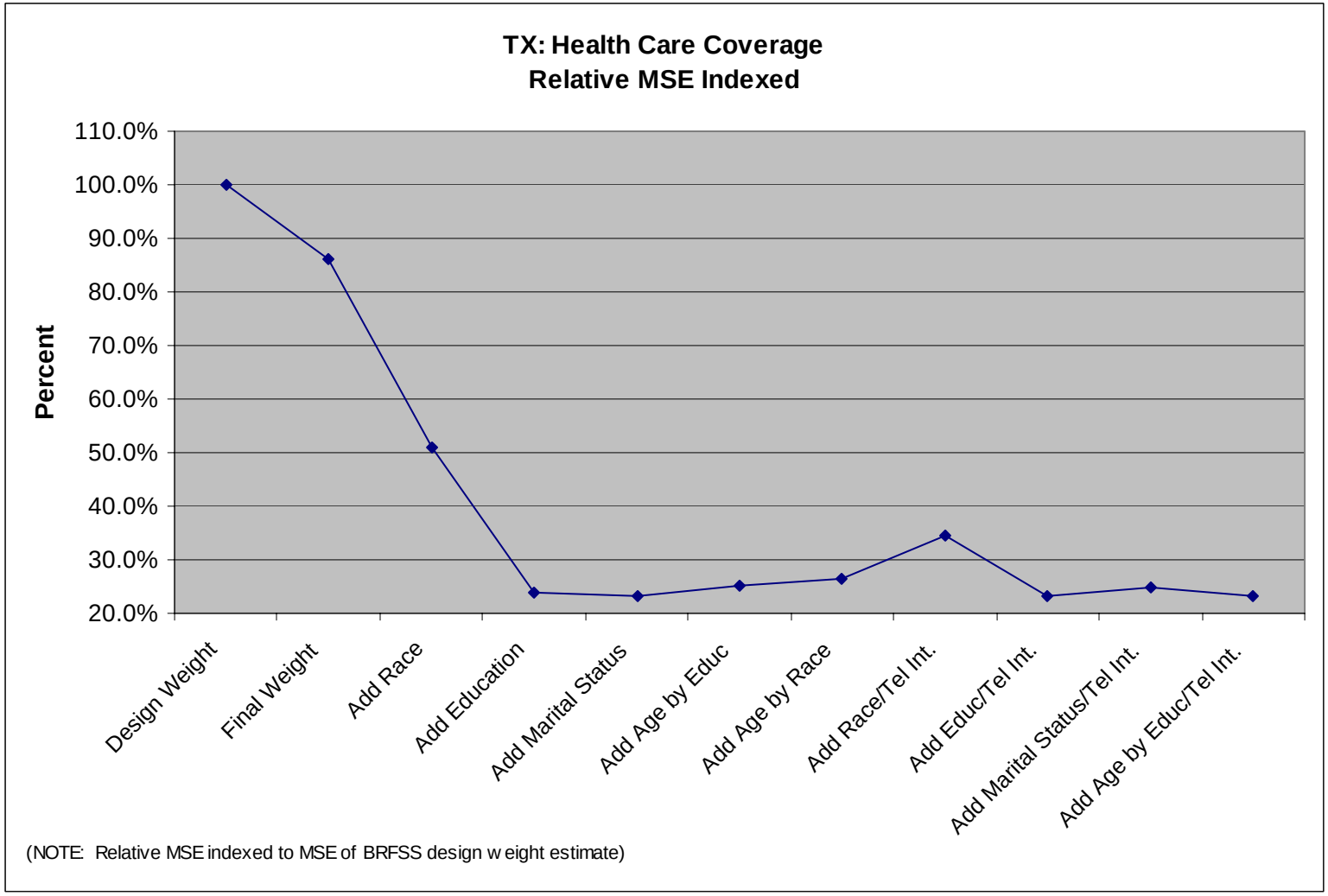

CA: Health Status
Relative MSE Indexed
Percent
120.0%
100.0%
80.0%
60.0%
40.0%
20.0%
0.0%
Design Weight
Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Add Race/Tel Int.
Add Educ/Tel Int.
Add Marital Status/Tel Int.
Add Age by Educ/Tel Int.
(NOTE: Relative MSE indexed to MSE of BRFSS design weight

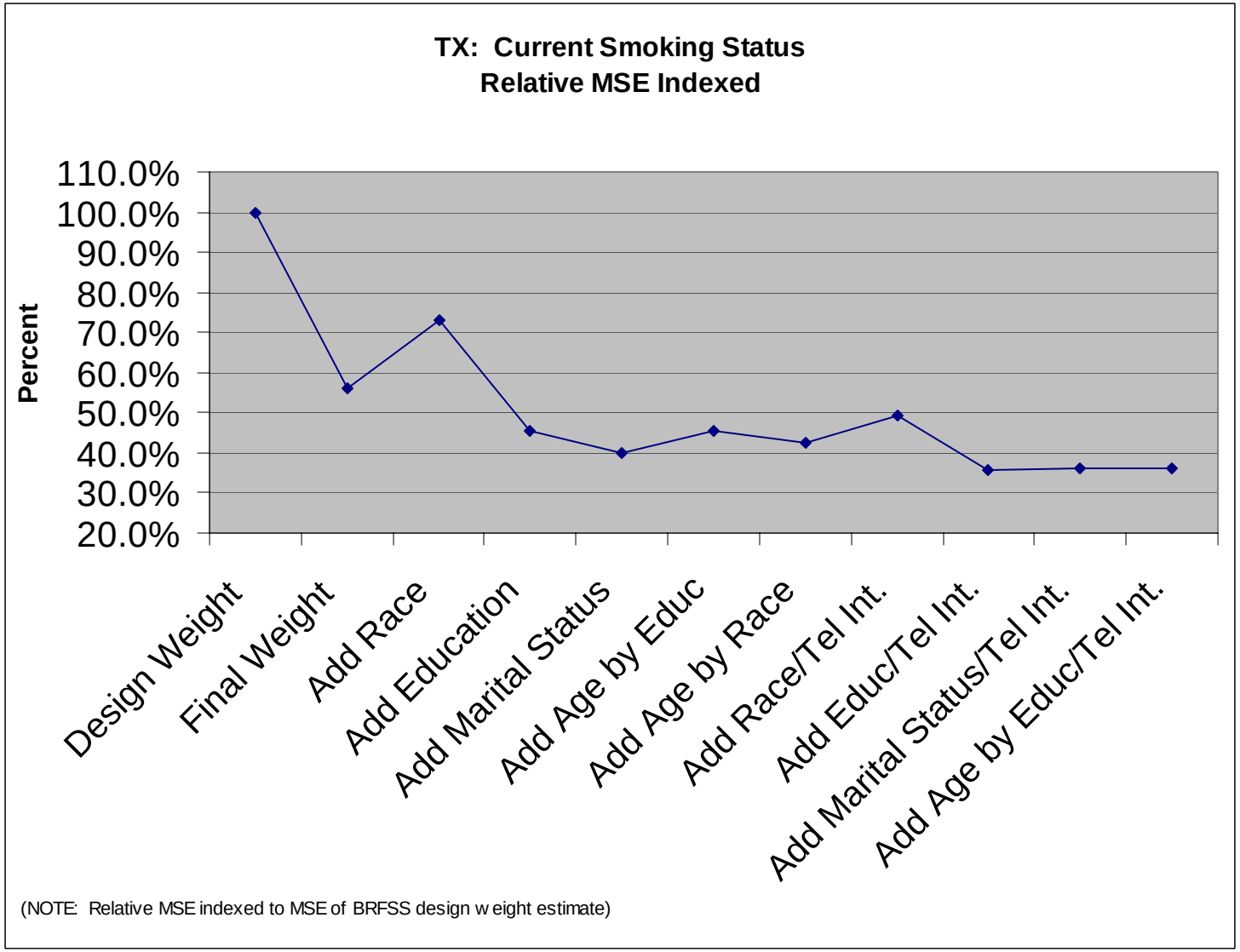


(NOTE: Relative MSE indexed to MSE of BRFSS design weight estimate)


CA: Smoking Status
Relative MSE Indexed
Percent
110.0%
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
Design Weight
Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Add Race/Tel Int.
Add Educ/Tel Int.
Add Marital Status/Tel Int.
Add Age by Educ/Tel Int.
(NOTE: Relative MSE indexed to MSE of BRFSS design weight estimate)


TX: Health Status
Relative MSE Indexed
Percent
140.0%
120.0%
100.0%
80.0%
60.0%
40.0%
20.0%
Design Weight
Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Add Race/Tel Int.
Add Educ/Tel Int.
Add Marital Status/Tel Int.
Add Age by Educ/Tel Int.
(NOTE: Relative MSE indexed to MSE of BRFSS design weight estimate)


TX: Health Care Coverage
Relative MSE Indexed
Percent
110.0%
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
Design Weight
Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Add Race/Tel Int.
Add Educ/Tel Int.
Add Marital Status/Tel Int.
Add Age by Educ/Tel Int.
(NOTE: Relative MSE indexed to MSE of BRFSS design weight estimate)


TX: Current Smoking Status
Relative MSE Indexed
Percent
110.0%
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
Design Weight
Final Weight
Add Race
Add Education
Add Marital Status
Add Age by Educ
Add Age by Race
Add Race/Tel Int.
Add Educ/Tel Int.
Add Marital Status/Tel Int.
Add Age by Educ/Tel Int.
(NOTE: Relative MSE indexed to MSE of BRFSS design weight estimate)

Conclusions

We have summarized the results from past research of identifying sociodemographic variables related to nonresponse in the 2003 BRFSS. We then illustrated the use of logistic regression and CHAID segmentation trees to identify sociodemographic variables associated with the 13 risk factor outcome variables in the BRFSS. It is important to focus on variables related to key survey outcome measures. Interestingly, for the 2003 BRFSS we found a fair amount of overlap between variable related to nonresponse and variables related to key survey outcome variables.

We then showed how to take existing age by gender or age by gender by race/ethnicity control totals (from Claritas) and develop revised CPS weights that are in agreement with those totals. We then used the revised CPS weight to develop control totals for the variables identified for inclusion in the rakings. In each raking we included a margin for the BRFSS age by gender or age by gender by race/ethnicity variable. This allowed us to "hold constant" the effect of including this variable in the weighting procedure. We also included a detailed race/ethnicity margin even if in a state a two-category race/ethnicity variable was used in the age by gender by race/ethnicity BRFSS margin. Two key findings emerged for the six states we examined: 1) the inclusion of additional variables in the raking raised many of the risk factor estimates, and 2) education is an important variable to include in the raking. In terms of the nontelephone adjustment using the interruption in telephone service approach, although it will typically cause an increase in the variance, for outcome variables associated with telephone status the adjustment can reduce noncoverage bias by a substantial amount. This is in line with the finding in Frankel et al. (2003), which used the National Health Interview Survey to assess the effectiveness of the interruption in telephone service adjustment.

Our next steps include running the raking for all 50 states and the District of Columbia, examining the need to trim high weights, and producing risk factor estimates for all states and DC combined and comparing those estimates with national risk factor estimates from the NHIS. This will provide a more direct way to assess bias reduction.

References

To be added.