Trends in Smoking Prevalence by Race based on the Tobacco Use Supplement to the Current Population Survey

William W. Davis[1], Anne M. Hartman[1], James T. Gibson[2]

National Cancer Institute, Bethesda, MD,[1] Information Management Systems, Silver Spring, MD[2]

Correspondence to: William W. Davis. Statistical Research and Applications Branch, Surveillance Research Program, National Cancer Institute. 6116 Executive Boulevard, Room 5047, Rockville, MD 20852. E-mail: davisbi@mail.nih.gov.

Abstract

The Tobacco Use Supplement to the Current Population Survey (TUS-CPS) can be used to obtain U.S. adult current smoking prevalence estimates for the recent past. In this paper, we use the TUS-CPS to evaluate current smoking prevalence rates and trends by race/ethnicity and gender for the period 1992 to 2003. The TUS-CPS adult sample size of approximately 240,000/year allows accurate estimates of current smoking rates for small race/ethnicity groups, which have often been ignored due to inadequate sample size. In this paper, we provide current smoking prevalence rates and trends for Hispanics and four non-Hispanic races (Whites, Blacks, Asian and Pacific Islanders (API), and American Indian and Alaskan Natives (AIAN)).

Due to a directive from the Office of Management and Budget (OMB), the Current Population Survey (CPS) changed its race/ethnicity questions in January 2003. In this paper, we utilized a previously developed race/bridging methodology to estimate the TUS-CPS current smoking prevalence trend for the 1992 to 2003 time period, which included survey results from before and after the wording change.

Statistically significant decreases in current smoking rates were obtained during this period for all race/ethnicity*gender combinations with the exception of the Asian and Pacific Islanders (API) females. Historically, the non-Hispanic AIAN group has had the highest current smoking rates for both genders. However, during the 1992-2003 time period, they showed the largest estimated decrease among the five race/ethnicity groups for both genders. The current smoking rates of the non-Hispanic AIAN were compared with the non-Hispanic White reference group both adjusting and not-adjusting for covariates; this comparison demonstrated the large impact that socio-economic status (SES) has on current smoking.

Keywords: disparity, ethnicity, multiple imputation, predictive margins

1. Introduction

Scientific knowledge about the health effects of tobacco use has increased greatly since the first Surgeon General's report on tobacco was released in 1964 (U.S. Department of Health and Human Services (DHHS), 2004). Tobacco use is responsible for more than 440,000 deaths per year among adults in the United States, representing more than 5.6 million years of potential life lost (Centers for Disease Control and Prevention, 2002).

The National Cancer Institute (NCI) as part of the U.S. DHHS is committed to reducing health disparities. An important goal of Healthy People 2010 is to help eliminate health disparities among different segments of the U.S. population (U.S. DHHS, 2000) including race/ethnicity groups. Tobacco related health disparities (TRHD) include differences in tobacco exposure and the subsequent health consequences among specific population groups such as those defined by race/ethnicity (Fagan *et al.*, 2007a). In this paper, we study current smoking prevalence by race/ethnicity as one important aspect of TRHD. In particular, we study whether the difference in current smoking prevalence rates, as defined by race/ethnic groups, is increasing or decreasing over the recent past. For previous studies of current smoking prevalence by race/ethnicity using the TUS-CPS see Shavers *et al.* (2005, 2006) and Fagan *et al.* (2007b).

Several national surveys have helped us to understand differences in smoking among racial groups. The National Health Interview Survey (NHIS) is used as the gold standard to determine trends in U.S. adult smoking prevalence. U.S. adult current smoking prevalence trends for the last four decades are available using NHIS responses (U.S. DHHS, 2007). The Tobacco Use Supplement to the Current Population Survey (TUS-CPS) also can provide U.S. adult current smoking

prevalence estimates. When compared with the NHIS, the TUS-CPS has a shorter history and is conducted less frequently. However, when it is conducted, the TUS-CPS has a much larger yearly sample size than the NHIS so that it allows more accurate estimates of smoking prevalence for small groups such as those defined by race and ethnicity.

Since some of the smaller race/ethnicity groups are underserved, we need to quantify health trends by race/ethnicity – including smoking trends. In this paper, we study current smoking prevalence and trends among U.S. adults for the period 1992-2003 by race/ethnicity and gender using information from the TUS-CPS. While current smoking percentage ignores important aspects such as the amount smoked, it is an extremely important tobacco-related health indicator and is the focus of this paper. The large yearly TUS-CPS adult sample size of approximately 240,000 allows accurate estimates of current smoking prevalence rates for smaller groups such as the American Indian and Alaskan Native (AIAN). The NHIS yearly sample size of approximately 40,000 adults (Botman *et al.*, 2000) is not sufficient to provide precise estimates for small subgroups such as the AIAN.

In order to study current smoking prevalence and trends for this time period, we must deal with a change in reporting of race/ethnicity. Due to a directive from the Office of Management and Budget (1997), the Current Population Survey (CPS) changed its race/ethnicity questions in January 2003. The major differences in the CPS race/ethnicity questions are the following:
1. Respondents may now select more than one race when answering the survey.
2. The Asian or Pacific Islander (API) category was split into two categories: Asian and Native Hawaiian or Other Pacific Islander (NHOPI).
3. The ethnicity question was reworded to ask directly whether the respondent was Hispanic (an ethnicity rather than a race).
4. The ethnicity and race questions order were reversed.

With the post-2003 race/ethnicity classifications, TUS-CPS data users can calculate descriptive statistics using "only" or "any mention" race/ethnicity categories. For example, the user can calculate current smoking prevalence rates of Non-Hispanic (NH) Black males 18+ including those who list "NH Black Only" or including those who list NH Black (and other races) for the "any mention" estimate. The "only" classification gives too little emphasis to the mixed race respondents while the "any mention" classification gives too much emphasis. Neither of these estimates is comparable to the pre-2003 TUS-CPS system, where multi-race reporting was not allowed. The new system can be used to begin current smoking trends for many new race/ethnic categories defined by multiple races. However, it is also important to continue the trends established using the pre-2003 race/ethnicity classifications. We do this applying a previously developed method.

In May 2002, the Bureau of Labor Statistics (BLS) sponsored a CPS supplement that asked the new (post-2003) race/ethnicity question to all sample people. The BLS used this sample to compare estimates of unemployment using the pre-2003 and the post-2003 race/ethnicity responses (Bowles *et al*, 2003, Tucker *et al*, 2002).

Using the information obtained from the Census Bureau from the May 2002 CPS supplement, Davis *et al* (2007) developed a race bridging approach that allows the user of post-2003 TUS-CPS data to calculate estimates that are comparable to the pre-2003 TUS-CPS estimates. The method uses the post-2003 race/ethnicity responses to multiply impute (MI) the (unknown) pre-2003 race/ethnicity response. The imputation method can be applied to a post-2003 dataset, which allows race/ethnicity estimates on the post-2003 dataset that are comparable to those that would have been obtained if the pre-2003 race/ethnicity questions had been used. This is called forward bridging since it allows trends established under the pre-2003 race/ethnicity questions to be bridged forward using the post-2003 race/ethnicity questions. There are only four possible imputed responses using the pre-2003 classification system: White, Black, AIAN, and API (and all are non-Hispanic -- so the method imputes race for non-Hispanics). The method imputes race for respondents who specify multiple races in the post-2003 system. Thus, the method is particularly useful for racial groups whose respondents often report multiple races. This is the case for two races that are often underserved; namely, the AIAN and the NHOPI.

In this paper, we emphasize prevalence and trends in current smoking by race/ethnicity for the period 1992 to 2003. In particular, we extend current smoking prevalence trends for Hispanics and four non-Hispanic races (Whites, Blacks, AIAN, and API). Since the NHOPI group was not specified prior to 2003 (see #2 above), we do not include them in the analyses of this paper explicitly; however, they are included in the analysis as a component of the API group. Due to their high rates, we emphasize the NH AIAN current smoking prevalence rates and trends, and determine if declines in their current smoking prevalence rates are larger or smaller than those obtained by other race/ethnicity groups during this period.

In Section 2, we describe the data and statistical methods used in the paper. In Section 3, we provide a comparison of the 2003 multiply imputed current smoking estimates with those obtained using the "only" and "any mention" classification by race/ethnicity and gender. We also provide estimates of current smoking prevalence trends for the period 1992-2003 by race/ethnicity and gender. We use the trend estimates to assess whether the race/ethnic smoking differences are increasing or decreasing. Also, to study the trend in the smoking disparity, we compare the non-Hispanic AIAN and non-Hispanic White current smoking prevalence rates by year – adjusting for the difference in socio-economic status (SES) and using the 2003 multiply imputed race/ethnicities. In Section 4, we discuss the implications of the results and suggest future research areas.

2. Data and Methods

2.1 TUS-CPS Data Utilized

The TUS-CPS is a survey of tobacco use that has been administered as part of the CPS since 1992-1993. The CPS is a continuing monthly survey conducted by the Census Bureau for the Bureau of Labor Statistics. Five TUS-CPS survey waves were used in this analysis; 1992-3, 1995-6, 1998-9, 2001-2, and 2003. The first four survey waves were conducted approximately every three years while the 2003 survey wave was a Special Topics questionnaire oriented toward tobacco cessation and was conducted approximately 18 months after the 2001-2002 survey.

Each of these five TUS-CPS survey waves was conducted as three monthly supplements to the CPS. The three months are typically chosen to be four months apart in order to obtain unique individuals from the CPS panel sample and also to cover small seasonal variations in smoking prevalence. Each TUS-CPS monthly sample is nationally representative of the civilian non-institutionalized population of the United States and weighted to represent the population for that year. When the three months are combined for any given wave their weights are divided by three so that the aggregate represents the population for that year. For simplicity, we will use the term "yearly" to represent a wave of 3 monthly surveys combined together. Each TUS-CPS wave yields approximately 240,000 respondents. About 25% of the TUS-CPS respondents complete the survey in-person while the remainder completes the survey by telephone. Additional information about the TUS-CPS survey including future survey plans and previous publications using the survey data is provided at National Cancer Institute (2007).

The TUS-CPS survey allows both self- and proxy responses for different questionnaire items. A self-respondent answered the question her/him self. Proxies were respondents who were sufficiently knowledgeable to respond to questions for the intended person. The analyses conducted here included the information from both self- and proxy respondents, who are at least 18 years old. The TUS-CPS response rate including both self- and proxy responses for the five yearly survey waves were 88.0%, 86.2%, 84.8%, 82.8%, and 83.0% respectively for those 18 years and older. The average of these five yearly response rates is 84.9%.

2.2 Measures and Statistical Weights Utilized

The primary outcome of this paper, current smoking, was obtained from the CPS-TUS. The standard definition of a current smoker was used; namely, anyone who had smoked 100 cigarettes and who was currently smoking every day or some days. The remaining measures used to stratify or to adjust current smoking analyses were obtained from the CPS core questionnaire responses. All analyses were stratified by gender. All measures thought to be related to current smoking prevalence were included. In particular, information from the CPS concerning socio-economic status (SES) was included. All measures were categorical. The following measures were used (with the number of levels in parentheses): age (4), education (4), Census region (4), race/ethnicity (5), survey year (5), metropolitan status (2), employment status (3), family income (3), and gender (2). The information for all these variables was complete with the exception of family income. Additional categories were added when necessary to reflect missing values so that all survey results could be utilized in the analysis.

Each of the five TUS-CPS survey waves contains a supplement non-response and a self-response survey weight. The non-response weight controls for supplement survey non-response and is used when both proxy- and self-reports (all respondents) are utilized. The self-response weight adjusts the self-respondents to appropriate demographic totals and is used for analysis using only self-respondents (excluding proxy respondents).

Although they are not distributed with the public use TUS-CPS file, replicate weights (both non-response and self-response) are available from NCI on request for each survey wave. These replicate weights have been derived using balanced repeated replication (BRR), and they can be used to provide estimates of variance that utilize the TUS-CPS survey design and response rates. The non-response replicate weights are used for analyses involving all respondents (self- and proxy) while self-response weights are used for analyses involving only self respondents. The number of replicate weights was 48 for the

1992-1993 survey and was 80 for the other four later surveys. For additional information about the CPS design and the statistical weights see Current Population Survey (2002).

2.3 Statistical Methods

Standard adjustments were made to the supplement and supplement replicate weights so that totals were averages of the monthly totals. For example, monthly weights were multiplied by 1/3 when using results from a single survey year (composed of 3 monthly surveys) and multiplied by 1/15 when using the five yearly surveys simultaneously.

The majority of the analyses of this paper were conducted for a single survey year. In this case, the weighted estimates are representative of the U.S. population for that year. However, to conserve space, the results of some analyses are shown using the pooled responses from all five survey waves. The weighted estimates from this pooled analysis can be considered as representative of an average U.S. population over this 11 year period. Since the U.S. population increased over the time period and three of the five surveys were conducted in the second half of the time period, the weighted results emphasize the later years in the period.

SUDAAN (Research Triangle Institute, 2004) software was used in the analyses in order to calculate standard errors and confidence intervals (CI) that are valid for the complex stratified CPS design. SUDAAN also allows use of the replicate weights. Fay's (1984) method for variance estimation was used. SUDAAN's implementation of Fay's method involves a parameter which was specified as ADJFAY=4 (e.g., Research Triangle Institute, 2004, Sec. 3.4.6).

Five multiply imputed race (MI) values were used for multiple race responders in the TUS-CPS 2003 survey (Rubin, 1987). As a check on the MI methodology, multiply imputed estimates of current smoking for 2003 were compared with the direct estimates using two different groups ("only" and "any mention") by race/ethnicity and gender. The Gelman and Rubin (1992) method was used to calculate the mean and variance of the five different estimates when using the MI race/ethnicities.

An ordinary least squares (OLS) regression was carried out to quantify trend in the current smoking prevalence rates over the time period by race/ethnicity and gender. To carry out the regression of the prevalence rates on time, the time of the survey wave was computed as the (midpoint of the) date of the second of the three component surveys. For example, the three 1992-1993 monthly survey were conducted in September 1992, January 1993, and May 1993. Thus, January 15th 1993 was considered the midpoint of the survey so for the 1992-1993 survey wave time was coded as 1993.04 (since on January 15th 1/24=0.04 of the year has been completed). In a sensitivity study, a weighted least squares (WLS) regression was carried out to estimate the robustness of the slope estimates, where the weights were chosen inversely proportional to the variance of the estimate of the prevalence estimate (Neter and Wasserman, 1974).

The current smoking prevalence race/ethnic differences over the time period were assessed in two ways. First, the sample range was computed and used to determine if the race/ethnic variation in smoking prevalence rates were increasing or decreasing over time. Second, the OLS slope estimates (along with the current smoking prevalence levels) were used to determine whether the current smoking prevalence rates were converging or diverging.

A logistic regression was developed for current smoking using covariates known to be related to smoking propensity such as demographic and SES measures (education, family income, employment status). All covariates were used as categorical in the logistic regression models. For ease of interpretation, the reference groups in the logistic regression were chosen so that the odds ratios of the other groups would be greater than 1.

Confidence intervals for the difference in current smoking rates between non-Hispanic (NH) AIANs and NH Whites were obtained by gender and year. The unadjusted and adjusted differences were contrasted. Predictive margins (Korn and Graubard, 1999) from a logistic regression equation were used to adjust for difference in covariate distributions between NH AIAN and NH Whites.

3. Results

3.1 Data summary

Table 1 shows sample counts, population size and percentage population for covariates and current smoking by gender. The table includes information about age, education, region, race/ethnicity, survey year, metropolitan status, employment status, and family income by gender. The table shows that about 10% of the income values were unknown, about 0.3% of the metropolitan statuses were not identified and 328 of the respondents were classified as Non-Hispanic Other.

Table 1. Sample size and weighted population size by gender for the aggregate of five TUS-CPS surveys conducted during the period 1992-2003

	Males			Females		
	Sample Size	Population	Percent †	Sample Size	Population	Percent †
Total	560,111	94,494,060	100.0	642,924	103,110,362	100.0
Current Smoker						
Yes	134,253	22,563,844	23.9	125,792	19,628,324	19.0
No	425,858	71,930,216	76.1	517,132	83,482,038	81.0
Age						
18-24	66,391	12,842,510	13.6	71,687	12,996,493	12.6
25-44	233,546	40,284,486	42.6	257,392	41,717,051	40.5
45-64	172,240	27,846,180	29.5	188,831	29,824,820	28.9
65+	87,934	13,520,885	14.3	125,014	18,571,999	18.0
Education						
< 12 years	95,120	16,677,408	17.6	107,295	17,769,869	17.2
12 years	180,808	29,820,647	31.6	221,501	34,736,696	33.7
13-15 years	142,152	24,160,631	25.6	175,740	28,263,131	27.4
16+ years	142,031	23,835,374	25.2	138,388	22,340,666	21.7
Region						
Northeast	123,330	18,283,511	19.3	144,472	20,429,905	19.8
Midwest	138,900	21,909,317	23.2	157,530	23,778,052	23.1
South	164,695	33,076,637	35.0	194,414	36,613,717	35.5
West	133,186	21,224,595	22.5	146,508	22,288,688	21.6
Race/Ethnicity						
Non-Hispanic White	445,544	70,279,932	74.4	500,433	75,815,871	73.5
Non-Hispanic Black	43,916	9,898,209	10.5	63,624	12,468,082	12.1
Non-Hispanic AIAN	5,740	647,296	0.7	6,661	717,616	0.7
Non-Hispanic API	19,011	3,434,482	3.6	22,075	3,854,492	3.7
Hispanic	45,739	10,212,020	10.8	49,963	10,233,953	9.9
Non-Hispanic Other ††	161	22,122	0.0	167	20,347	0.0
Survey						
1992-93	127,377	88,350,523	18.7	148,518	96,991,062	18.8
1995-96	107,524	91,208,453	19.3	126,213	99,866,493	19.4
1998-99	105,176	94,369,167	20.0	119,726	102,996,253	20.0
2001-02	109,993	97,298,124	20.6	124,234	105,814,645	20.5
2003	110,041	101,244,033	21.4	124,233	109,883,356	21.3
Metropolitan Status						
Metropolitan	412,452	75,736,094	80.2	477,294	82,869,291	80.4
Non Metropolitan	143,832	18,437,043	19.5	161,401	19,905,624	19.3
Not Identified ††	3,827	320,923	0.3	4,229	335,468	0.3
Employment Status						
Employed	402,843	68,309,993	72.3	366,939	58,965,279	57.2
Unemployed	22,953	4,125,425	4.4	20,064	3,458,146	3.3
Not in Labor Force	134,315	22,058,642	23.3	255,921	40,686,937	39.5
Family Income						
< $25,000	147,817	24,461,103	25.9	208,691	32,506,610	31.5
$25,000 - $49,999	165,656	27,103,218	28.7	178,298	27,895,534	27.1
$50,000 or more	194,650	33,692,707	35.7	192,783	32,047,369	31.1
Unknown	51,988	9,237,032	9.8	63,152	10,660,900	10.3

† Column may not sum to 100% due to rounding

†† All values were obtained from the 1992/1993 survey

With the exception of survey year, the populations (and percentage population) in table 1 are weighted and reflect an average U.S. population size (age 18+) over the eleven year time period 1992-2003 so they sum to the population total (modulo rounding differences). For the survey year, the population reflects the total for the survey wave. The population estimates for race/ethnicity were computed as an average over the five multiply imputed estimates.

Table 1 shows that there were over 1.2 million respondents to the TUS-CPS during this time period including more than 560,000 males and 642,000 females. Inspection of the table shows that there were 224,000 or more responses in every survey year. The large TUS-CPS sample allows analysis of subgroups that are a small fraction of the U.S. population. Due to their high current smoking prevalence rates, we emphasize results for the Non-Hispanic American Indian and Alaskan Natives in this paper. Table 1 shows that even though this group represents less than 1% of the U.S. population, over 5,700 non-Hispanic AIAN males and over 6,600 non-Hispanic AIAN females responded to the TUS-CPS survey during this time period.

3.2 Comparison of current smoking estimates for 2003 using three different group classifications

Table 2 compares the current smoking prevalence estimates obtained from TUS-CPS 2003 by three different group classifications ("any mention", "only", and MI) by race/ethnicity and gender. For these classifications, the table shows the current smoking prevalence estimate and the 95% CIs. Due to the method of group classification, the Hispanic estimates are identical. Also, the MI estimates are not available (NA) for three non-Hispanic groups (Asian, NHOPI and multiple) since these groups were not identified until 2003. For both genders, table 2 shows the highest current smoking prevalence rates for the NH AIAN followed by the NH multiple group responders.

The table also shows the number of only (single race) mention and any mentions by race/ethnicity and gender; in addition, it shows the ratio of the number only/any mentions. If this ratio is near 100% for a group, the current smoking prevalence estimates for the three classifications of table 2 must be similar (since there are relatively few multi-race responders). This is the case for NH White, NH Black, and NH Asian where the ratios are over 90%. However, for the NH AIAN, the ratio is less than 50% for both genders indicating a large percentage of mixed-race responders. For this group, the current smoking prevalence estimates could vary considerably between the three classifications.

Table 2. Current Smoking Estimates Using TUS-CPS 2003

| Race/ethnicity | Only Mention | | Any Mention | | Ratio Only/Any | Multiple Imputation |
	Number	Estimate (95% CI)	Number	Estimate (95% CI)		Estimate (95% CI)
Females						
Total	124,233	16.24 (15.93, 16.56)	124,233	16.24 (15.93, 16.56)	100.0%	16.24 (15.93, 16.56)
NH White	93,962	18.04 (17.68, 18.42)	95,377	18.16 (17.80, 18.53)	98.5%	18.14 (17.78, 18.51)
NH Black	11,770	15.35 (14.50, 16.24)	12,075	15.47 (14.63, 16.34)	97.5%	15.39 (14.55, 16.27)
NH AIAN	1,049	30.53 (26.17, 35.27)	2,198	28.39 (25.99, 30.93)	47.7%	29.97 (26.10, 34.15)
NH API	4,463	4.72 (3.97, 5.59)	4,796	5.31 (4.54, 6.20)	93.1%	5.14 (4.37, 6.03)
NH Asian	4,141	4.35 (3.62, 5.21)	4,387	4.81 (4.05, 5.71)	94.4%	NA
NH NHOPI	322	12.51 (8.74, 17.60)	489	16.14 (12.22, 21.01)	65.8%	NA
Hispanic	11,308	9.11 (8.47, 9.80)	11,308	9.11 (8.47, 9.80)	100.0%	9.11 (8.47, 9.80)
NH Multiple	1,681	25.74 (23.23, 28.42)	1,681	25.74 (23.23, 28.42)	100.0%	NA
Males						
Total	110,041	20.69 (20.33, 21.07)	110,041	20.69 (20.33, 21.07)	100.0%	20.69 (20.33, 21.07)
NH White	84,730	21.38 (20.98, 21.78)	85,969	21.53 (21.14, 21.93)	98.6%	21.49 (21.10, 21.89)
NH Black	8,238	21.78 (20.83, 22.75)	8,473	21.81 (20.89, 22.76)	97.2%	21.80 (20.86, 22.77)
NH AIAN	969	32.83 (27.55, 38.59)	1,966	33.37 (30.30, 36.59)	49.3%	33.67 (28.47, 39.29)
NH API	3,771	16.59 (15.04, 18.26)	4,095	17.12 (15.59, 18.77)	92.1%	16.97 (15.43, 18.62)
NH Asian	3,511	16.71 (15.14, 18.40)	3,772	17.23 (15.68, 18.90)	93.1%	NA
NH NHOPI	260	13.79 (9.50, 19.60)	419	16.27 (11.78, 22.04)	62.1%	NA
Hispanic	10,833	16.08 (15.29, 16.90)	10,833	16.08 (15.29, 16.90)	100.0%	16.08 (15.29, 16.90)
NH Multiple	1,500	32.08 (29.29, 35.00)	1,500	32.08 (29.29, 35.00)	100.0%	NA

* NA – not available due to TUS-CPS race/ethnicity data limitation prior to 2003

Figure 1 compares the three (any mention, only, and MI) estimates (and 95% CIs) of current smoking for the NH American Indian and Alaskan Natives by gender. The figure shows that for both genders the point estimates are similar while the confidence intervals are shorter for the "any mention" classification – due to the larger number of respondents in the group.
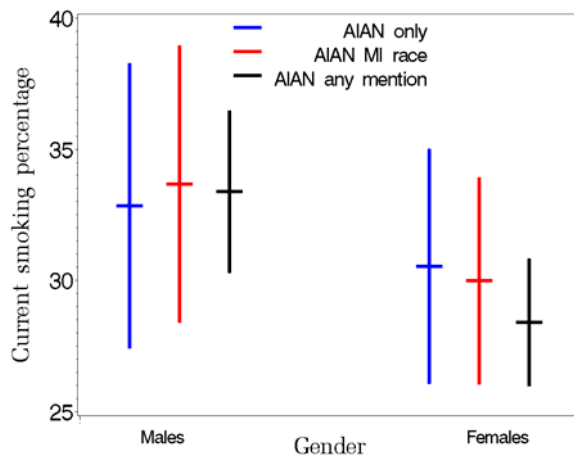

Figure 1. Current smoking prevalence estimates (with 95% CI) by gender from the TUS-CPS 2003 for non-Hispanic (NH) American Indian and Alaskan Native (AIAN) only, NH AIAN any mention, and NH AIAN with multiply imputed (MI) race

3.3 Trend in Current Smoking from 1992 to 2003 by Race/ethnicity and gender using TUS-CPS
Table 3 shows current smoking prevalence estimates and 95% confidence intervals by race/ethnicity and gender for the five TUS-CPS survey waves carried out in the period 1992-2003. For 2003, the multiply imputed current smoking prevalence estimates from Table 2 are used.

Table 3. Current smoking prevalence rate estimates (with 95% CI) by race/ethnicity and gender from TUS-CPS: 1992-2003

	1992-1993	1995-1996	1999	2001	2003
Females					
All	21.45	20.73	19.05	18.11	16.24
	(21.14, 21.77)	(20.45, 21.01)	(19.80,19.30)	(17.87,18.36)	(15.93,16.56)
NH White	22.80	22.51	20.78	19.99	18.14
	(22.45, 23.16)	(22.19, 22.83)	(20.52,21.05)	(19.69,20.30)	(17.78,18.51)
NH Black	21.94	19.56	18.20	17.05	15.39
	(21.13, 22.77)	(18.41, 19.92)	(17.36,19.07)	(16.17,17.96)	(14.55,16.27)
NH AIAN	35.96	35.43	31.29	29.99	29.97
	(31.35, 40.84)	(31.05, 40.06)	(27.64,35.19)	(26.33,33.91)	(26.10,34.15)
NH API	7.87	6.83	7.27	6.67	5.14
	(7.08, 8.75)	(5.94, 7.84)	(6.47,8.17)	(5.87,7.57)	(4.37,6.03)
Hispanic	12.28	12.08	10.67	9.96	9.11
	(11.59, 13.01)	(11.32, 12.88)	(9.94,11.46)	(9.28,10.68)	(8.47,9.80)
Males					
All	26.47	25.66	24.04	23.01	20.69
	(26.13, 26.81)	(25.32, 26.01)	(23.73,24.35)	(22.63,23.40)	(20.33,21.07)
NH White	26.31	25.73	24,44	23.49	21.49
	(25.97, 26.65)	(25.37, 26.09)	(24.05,24.82)	(23.08,23.90)	(21.10,21.89)
NH Black	30.16	27.91	25.80	24.21	21.80
	(29.04, 31.31)	(26.74, 29.11)	(24.60,27.04)	(23.13,25.33)	(20.86,22.77)
NH AIAN	42.26	41.76	35.73	34.78	33.67
	(37.31, 47.37)	(37.56, 46.08)	(32.88,38.67)	(30.76,39.03)	(28.47,39.29)
NH API	22.11	21.91	18.86	17.72	16.97
	(20.19, 24.16)	(20.06, 23.89)	(17.35,20.47)	(16.24,19.30)	(15.43,18.62)
Hispanic	24.12	22.85	20.47	19.72	16.08
	(22.94, 25.33)	(21.87, 23.87)	(19.69,21.28)	(18.82,20.66)	(15.29,16.90)

To measure whether the group differences (disparities) have increased or decreased over the 11 year time period, first we use a simple method based on the group range (= the largest value minus the smallest value), which is a measure of group variation. If the range increases (decreases) over time, we say that the group differences (disparities) have increased (decreased). From table 4, for females the range (over the five race/ethnic groups) decreases from 28.09 (= 35.96 - 7.87) in 1992/1993 to 24.83 (= 29.97 - 5.14) in 2003 while for males the range decreases from 20.15 (= 42.26 - 22.11) in 1992/1993 to 17.59 (= 33.67 - 16.08) in 2003. Thus, the range decreases for both genders indicating a reduction in current smoking prevalence disparities over the five race/ethnic groups.

Figure 2 show the current smoking estimates and 95% confidence intervals (CIs) by gender for five survey waves by race/ethnicity (jittered at each time period to avoid the overlap of confidence intervals). These figures indicate a decreasing time trend for all race/ethnicities and genders with the highest current smoking rates for NH AIAN for both genders. The figures also show higher current smoking rates for NH Blacks and NH Whites than for the NH API and the Hispanics for both genders. In both figures, the length of the CIs range from very narrow (NH White) to wide (NH AIAN) - reflecting the difference in the number of responses for these race/ethnicities by wave.
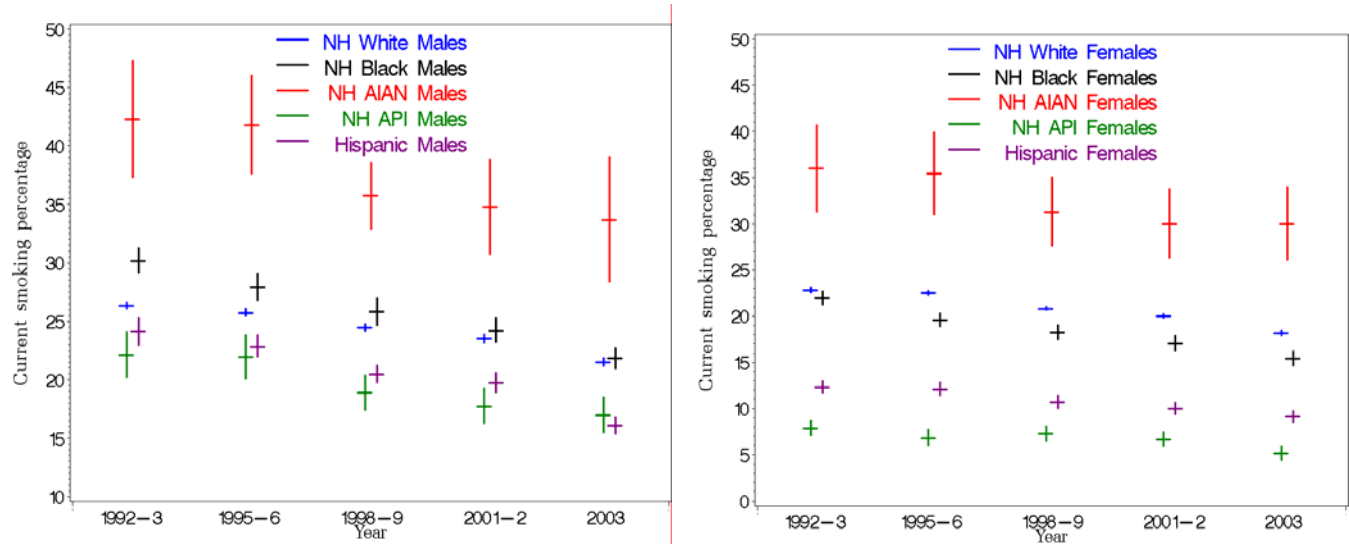

Figure 2. Current smoking prevalence rate estimates (with 95% confidence intervals) by race/ethnicity, year, and gender from TUS-CPS for the period 1992 to 2003

To summarize the change over time of the current smoking prevalence estimates by race/ethnicity and gender shown in Figure 2, we estimated the parameters of the linear model

$$r_t = \alpha + \beta t + e_t \tag{1}$$

where t is time, r_t denotes the estimated current smoking prevalence rate (from Table 3) at time t, and α and β denote the intercept and slope of the line respectively; the errors, e_t, are assumed to have mean 0 and variance σ^2. The parameters of (1) were estimated using ordinary least squares (OLS), and the estimate of trend, $\hat{\beta}$, is used as a summary measure.

Table 4 shows the slope estimate, $\hat{\beta}$, and its standard error by race/ethnicity and gender. The table shows a negative slope estimate for all race/ethnicities and genders – indicating a decrease in current smoking prevalence rates. The p-values of the test of non-zero slope (results not shown) show that, with the exception of API females, all slope decreases are statistically significant at the 5% significance level. For both genders including all race/ethnicities, the table indicates a decrease in current smoking of approximately 0.5%/year – yielding an approximate 5.5% decrease in current smoking prevalence over the eleven year period for the U.S. population. Figure 2 shows the highest current smoking levels for the NH AIAN throughout the period for both genders. However, table 4 shows the yearly decrease (as measured through the slope estimate) is greater for the NH AIAN than for any other race/ethnicity for both genders. This comparison of slopes implies a reduction in current smoking differences (disparity) between the NH AIAN and the other race/ethnicities.

In a sensitivity study, equation (1) was estimated using weighted least squares, with weights chosen inversely proportional to the error variance. The difference of the WLS estimates of slope from the OLS estimates of table 4 was small - so these results are not included.

Table 4. Estimated change in yearly current smoking prevalence rate by race/ethnicity and gender: obtained as ordinary least squares (OLS) estimate of slope for the period 1992-2003 using TUS-CPS data*

	Females		Males	
Race/ethnicity	Estimate (%)	Standard Error (%)	Estimate (%)	Standard Error (%)
All	-0.48	0.06	-0.52	0.08
NH White	-0.43	0.07	-0.43	0.08
NH Black	-0.58	0.05	-0.76	0.06
NH AIAN	-0.66	0.12	-0.91	0.16
NH API	-0.20	0.08	-0.54	0.08
Hispanic	-0.31	0.04	-0.70	0.13

* Slope estimates obtained from equation (1) using current smoking prevalence rates from Table 3

3.4 Change in Current Smoking from 1992 to 2003 by Race/ethnicity and gender using TUS-CPS
Figure 3 shows the difference in current smoking estimates (and the 95% confidence interval) between the last (2003) and first (1992-1993) TUS-CPS survey waves (thus ignoring the information from the three middle surveys). Since the current smoking prevalence estimates decreased systematically during this period, the difference estimates shown in the figure are all negative. With the exception of the NH AIAN females, all of the 95% confidence intervals do not overlap zero indicating a statistically significant decrease (at significance level 0.05) over the time period. For both genders, the largest estimated decreases are obtained for the NH AIAN and the NH Blacks, which is consistent with the Section 3.3 conclusion (see table 4) that these racial groups have the largest negative trend estimates.
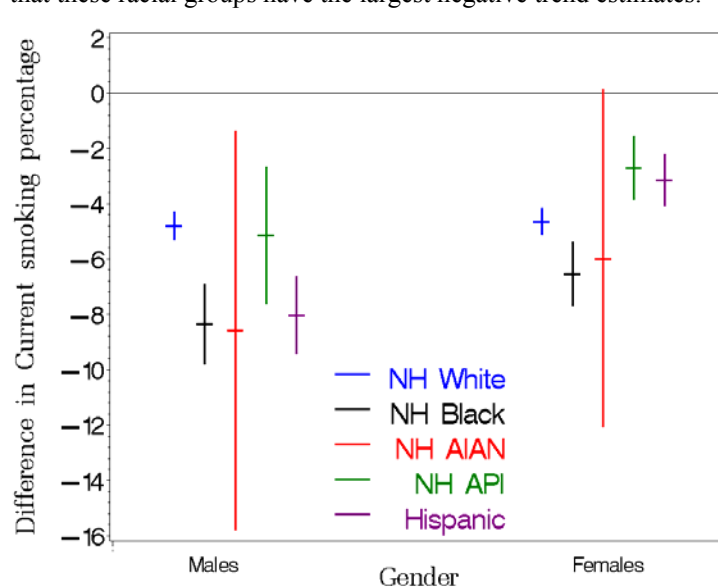

Figure 3. Estimate (and 95% confidence interval) for the difference in the current smoking prevalence rate between 1992-1993 and 2003 (negative values indicate a decrease) by race-ethnicity and gender using TUS-CPS data

3.5 Logistic regression results for the 1992-2003 time period using TUS-CPS
 To show the impact of the covariates of table 1 on the current smoking prevalence rates, a multiple logistic regression equation of current smoking status was carried out by gender. The odds-ratios (ORs,) and their 95% confidence interval obtained from the multiple logistic regression are shown in Table 5. Although variation of the ORs by survey wave may be of interest, the OR estimates are shown for the five pooled survey waves to save space. In order to assess the potential for bias due to the use of proxy-responses for smoking status, we also included respondent status (self, proxy, and unknown) in the logistic regression equation. The proportion of self responses was 73.6% for males and 84.9% for females while the percentage of unknown response status was very small (0.02% for both males and females - all from the 1992-1993 survey).

Table 5. Logistic regression results for current smoking prevalence by gender using results from the TUS-CPS for 1992-2003, entries are odds ratios (and 95% confidence intervals)*

	Males	Females
Age	Odds Ratio (95% CI)	Odds Ratio (95% CI)
18-24	2.83 (2.72, 2.96)	3.27 (3.15, 3.39)
25-44	4.10 (3.95, 4.26)	4.57 (4.43, 4.72)
45-64	3.63 (3.50, 3.76)	3.41 (3.31, 3.52)
65+	1.00	1.00
Education		
< 12 years	4.02 (3.89, 4.14)	3.84 (3.72, 3.95)
12 years (HS)	3.16 (3.08, 3.25)	3.08 (2.99, 3.16)
13-15 years	2.23 (2.17, 2.29)	2.29 (2.23, 2.35)
16+ years	1.00	1.00
Region		
Northeast	1.05 (1.02, 1.08)	1.13 (1.10, 1.17)
Midwest	1.19 (1.16, 1.22)	1.21 (1.18, 1.24)
South	1.19 (1.16, 1.22)	1.14 (1.11, 1.17)
West	1.00	1.00
Survey		
1992-93	1.13 (1.10, 1.16)	1.08 (1.05, 1.11)
1995-96	1.16 (1.13, 1.20)	1.13 (1.10, 1.16)
1998-99	1.14 (1.11, 1.17)	1.10 (1.07, 1.13)
2001-02	1.13 (1.10, 1.16)	1.10 (1.08, 1.13)
2003	1.00	1.00
Race/ethnicity		
Non-Hispanic White	2.14 (2.08, 2.20)	3.65 (3.48, 3.82)
Non-Hispanic Black	1.58 (1.52, 1.64)	2.05 (1.94, 2.17)
Non-Hispanic AIAN	2.80 (2.51, 3.12)	4.82 (4.29, 5.40)
Non-Hispanic API	1.82 (1.71, 1.93)	1.02 (0.94, 1.10)
Non-Hispanic Other	1.27 (0.77, 2.09)	1.26 (0.70, 2.26)
Hispanic	1.00	1.00
Metropolitan Status		
Metropolitan	1.08 (0.96, 1.20)	0,97 (0.89, 1.07)
Non Metropolitan	1.02 (0.91, 1.13)	0.90 (0.82, 0.98)
Not Identified	1.00	1.00
Employment Status		
Employed	1.10 (1.07, 1.13)	1.18 (1.15, 1.20)
Unemployed	1.68 (1.61, 1.76)	1.82 (1.75, 1.89)
Not in Labor Force	1.00	1.00
Family Income		
Unknown	1.18 (1.14, 1.21)	1.25 (1.20, 1.29)
< $25,000	1.94 (1.89, 1.99)	2.01 (1.95, 2.06)
$25,000 - $49,999	1.41 (1.38, 1.44)	1.44 (1.40, 1.48)
$50,000 or more	1.00	1.00
Respondent		
Unknown	0.40 (0.21, 0.75)	0.26 (0.11, 0.62)
Proxy	0.92 (0.91, 0.94)	0.78 (0.76, 0.80)
Self	1.00	1.00

* Each of the predictor variables is statistically significant at level 0.01 for both genders as measured by the Wald F test

Due partially to the large sample size, all of the covariates were statistically significant at the 0.01 level using the p-value computed from the Wald F test for both genders in table 5 (p-values not shown). Since we typically picked the reference group as the level with the lowest smoking prevalence, variables that have groups with the largest odds ratios are interpreted as the most important predictors of current smoking prevalence. Table 5 shows that, for the most part, the OR estimates are similar for both genders with the following important predictors:

- Age is an important predictor for both genders with ORs greater than 2.7 for all age groups younger than 65+. This could reflect the change in smoking with age and could also reflect a cohort effect (a change in the U.S. smoking patterns over time).
- Education is also an important predictor with all ORs greater than 2.2 for all other education categories compared to the reference group of college graduates (or more).
- A number of large ORs were obtained for race/ethnicity when compared with the Hispanic reference group. For NH White, NH AIAN and NH Black the ORs were significantly higher for females than males while the reverse was true for NH API.
- Family income is an important predictor with ORs of approximately 2 for the lowest income group (<$25,000/year) compared to the highest income group (>$50,000/year).

The table shows the strong impact of SES on current smoking. Education is one of the most important predictors while income and employment status both show that higher SES individuals have lower smoking rates controlling for other factors. In contrast to the other covariates, region, metropolitan status, and survey year show smaller differences in odds ratios; in fact, the ORs for all levels of these factors range between 0.90 and 1.21 for both genders.

The table shows the potential for a small bias through the use of proxy responses in the estimation of current smoking prevalence. With self-response as the reference category the odds ratio for proxy-response was 0.92 for males and 0.78 for females – adjusting for other covariates. This indicates the possibility that the use of proxy-responses could cause a small under-estimate of the true current smoking prevalence (obtained using self-responses only).

3.6 Comparison of current smoking rates for NH AIAN with NH White by gender over the period

Due to their high levels, we provide further analysis of the current smoking prevalence rates by non-Hispanic AIANs. In this section, we compare the difference in current smoking rates between the NH AIANs and the majority, NH White group by gender and by survey wave. Differences (and 95% confidence intervals) between NH AIAN and NH White were calculated and shown in Table 6 -- both unadjusted, d_u, and adjusted, d_a. The adjusted differences were obtained using predicted margins in a logistic regression equation including the covariates shown in Table 5. The percentage reduction due to the covariates is defined by $p_r = 100\% * (1 - d_a / d_u)$. A 100% reduction is obtained when the two groups have the same mean value after adjusting for covariates ($d_a = 0$). In contrast, a 0% reduction is obtained when the adjusted and unadjusted mean values coincide ($d_a = d_u$) – implying no impact of the covariates. Table 6 shows that although the unadjusted mean differences decrease over the time period, they are more than 10 percentage points for both genders for all years. Also, the table shows that the percentage reduction in the mean difference varies from 53 to 74% – indicating that the difference in covariates explains a majority of the difference in current smoking prevalence rates between NH Whites and NH AIANs.

Figure 4 shows adjusted and unadjusted 95% confidence intervals for both genders by survey year. For each year, the figure shows similar results by gender but quite different results for adjustment type; in fact,

- There is considerable overlap for the confidence intervals for the two genders for both adjusted and unadjusted. however;
- There is not much overlap between the adjusted and unadjusted confidence intervals for each gender.

The figure also shows that three of the ten adjusted confidence intervals overlap zero; for these three cases we would not reject the hypothesis of a difference in the current smoking rates (controlling for the difference in covariates between the two race/ethnicities).

In summary, there are substantial current smoking prevalence differences between the two (NH White and NH AIAN) race/ethnicities, but differences in covariates explain a majority of the mean prevalence difference. Table 5 suggests that SES covariates such as education, employment status, and family income are important explanatory factors of current smoking prevalence.

Table 6. Difference in current smoking rate for non-Hispanic American Indian and Alaskan Native (AIAN) and non-Hispanic Whites by gender and survey year: adjusted and unadjusted for covariates

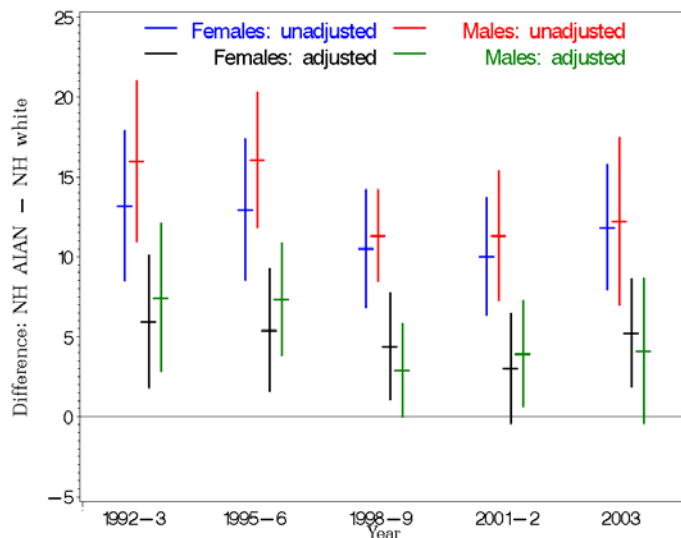|	1992-1993	1995-1996	1998-1999	2001-2002	2003
Females					
Unadjusted	13.16	12.92	10.51	10.00	11.83
(95% CI)	(8.42, 17.90)	(8.46, 17.38)	(6.79, 14.23)	(6.28, 13.72)	(7.89, 15.77)
Adjusted	5.92	5.40	4.35	3.00	5.23
(95% CI)	(1.74, 10.10)	(1.54, 9.26)	(0.97, 7.73)	(-0.50, 6.50)	(1.81, 8.65)
Percentage Reduction	55.0%	58.2 %	58.6%	70.0%	55.8%
Males					
Unadjusted	15.95	16.03	11.29	11.29	12.18
(95% CI)	(10.86, 21.04)	(11.75, 20.31)	(8.36, 14.22)	(7.19, 15.39)	(6.89, 17.47)
Adjusted	7.43	7.33	2.90	3.92	4.09
(95% CI)	(2.77, 12.09)	(3.77, 10.89)	(-0.66, 5.87)	(0.56, 7.28)	(-0.51, 8.69)
Percentage Reduction	53.4%	54.3%	74.3%	65.3%	66.4%

Figure 4. Difference in current smoking prevalence rate (in percent) for non-Hispanic American Indian and Alaskan Native (AIAN) and non-Hispanic Whites by gender and survey year: adjusted and unadjusted for covariates using TUS-CPS data

We chose to compare the smoking prevalence of the NH AIAN with the majority NH White reference group. If we had chosen a lower current smoking prevalence group such as Hispanics or NH API, we would have obtained even larger unadjusted differences with the NH AIANs for both genders and all surveys (e.g., Table 3).

4. Discussion

Although the National Health Interview Survey (NHIS) is used as the gold standard for determine trends in U.S. adult smoking prevalence, the Tobacco Use Supplement to the Current Population Survey (TUS-CPS) also furnishes current smoking prevalence estimates for the recent past. In this paper we emphasize U.S. adult current smoking prevalence and trend estimates using the TUS-CPS for the 1992 to 2003 time period. The large yearly TUS-CPS sample size allows us to study the prevalence and trend in current smoking prevalence rates by race/ethnicity and gender.

Due to the previously documented large impact of smoking on numerous health outcomes, smoking differences in subgroups can lead to outcome differences (such as incidence, mortality, and survival) for a wide variety of diseases. Among the

race/ethnicity subgroups studied here the Non-Hispanic American Indian and Alaskan Natives (NH AIAN) have the highest current smoking prevalence for both genders. It is well known the AIAN group has high smoking rates with large variation among tribes (U.S. DHHS, 1998). However, the trend analysis for the 1992-2003 period suggests that the NH AIAN had the largest decrease in current smoking -- as measured by the estimated slope of the OLS regression. If this trend continues, the current smoking prevalence of the NH AIAN will become more like those of the other race/ethnicity groups so that the race/ethnic smoking disparity of this group will be reduced relative to the rest of the population.

Harper and Lynch (2005) provide a number of alternative statistical analyses to determine whether the disparity (difference) in an outcome over multiple mutually exclusive groups is increasing (or decreasing) over time. They emphasize that different conclusions can be reached depending on the statistic chosen. An important statistical consideration is whether the groups are weighted by population or unweighted. A weighted analysis emphasizes the majority NH White values while an unweighted analysis gives equal weight to all groups – thereby overweighting the minority groups relative to their proportion of the population. The measures that we use (group range over time and comparison of group slope estimates) do not weight by race/ethnic population.

Harper and Lynch (2005, p. 17) discuss a number of different definitions of health disparities. All of these definitions emphasize the difference in an outcome measure (disease incidence, prevalence, morbidity, mortality, or survival rates) between two (or multiple) groups or between a specified group and the general population. For statistical purposes, it is easier to determine differences between two mutually exclusive groups than to determine the difference of a group from the general population. Thus, to determine the disparity of the current smoking prevalence for NH American Indian and Alaskan Natives, we compare their smoking prevalence with that of the NH White group, which is the dominant proportion of the U.S. population. We found substantial differences between the unadjusted and adjusted mean current smoking prevalence difference, which suggests a significant impact of the covariates. For an alternative method to determine the impact of the covariates on mean prevalence for two groups see Graubard *et al.* (2005).

The race-bridging multiple imputation (MI) method used in this study has some advantages over the "any mention" and "only mention" approaches. While figure 1 did not show substantial difference in the current smoking prevalence estimates among the three methods, MI offers advantages for complex analysis such as regression, where it is difficult, if not impossible, to use the other two approaches. Also, the race bridging MI approach allows pre- and post-2003 race/ethnicity results to be compared; no comparable pre-2003 results are available for the other two approaches.

4.1 Study Limitations and Future Research
The results of this paper were based on all respondents since the proportion of self responses (of the total number of respondents) was substantial (73.6% for males and 84.9% for females). Previous research suggests that there is a slight difference in current smoking prevalence estimates obtained through the use of proxy responses. It would be useful to repeat the race specific analyses of this paper using only self respondents to determine if the conclusions hold. For additional studies on the validity of smoking responses see Patrick *et al.* (1994), Gilpin *et al.* (1994), Hyland *et al.* (1997) and Caraballo *et al.* (2001).

This paper showed a decrease in current smoking prevalence rate disparity using two different analyses (based on the range and the slope estimates). It would be useful to determine if similar conclusions hold using the statistical measures and analyses that were proposed by Harper and Lynch (2005) and by Graubard *et al.* (2005). In addition, all the analyses of this paper were based on one measure of tobacco usage – current smoking prevalence. It would be useful to determine if a reduction in disparity is obtained for other measures of TRHD such as amount smoked, duration of smoking, age of initiation, quit rates, etc. (Levy *et al*, 2005).

Some of the study conclusions depend on the accuracy of the race/bridging method and multiply imputed race/ethnicities. The race/bridging method potentially has the largest impact on the current smoking prevalence estimates for races that have a large proportion of multi-race responders including the AIAN, whose results we emphasize here. The race bridging method was based on the May 2002 CPS overlap sample; the method would be more accurate if it were based on a larger sample. Also, since the race-ethnicity composition of the United States is changing constantly, the relationship between the post-2003 and the pre-2003 race/ethnicities, which was estimated in the race/bridging methodology based on a May 2002 sample, will eventually become inaccurate and will need to be periodically repeated. However, the application of the May 2002 overlap sample results to TUS-CPS samples obtained in 2003 (within 18 months of the overlap sample) should not suffer appreciably from this time degradation.

We used multiple imputation (MI) to allow previously established trends to be carried forward to the post-2003 CPS race/ethnicity questions. However, the post-2003 race/ethnicity responses can be used to study race/ethnicity groups other than the five studied here. The large yearly TUS-CPS sample size allows accurate estimates for a number of multiple race groups. It would be useful to use TUS-CPS to study the largest multiple race responder groups with respect to current smoking and other measures of TRHD; for example, the subset of non-Hispanic American Indians and Alaskan Natives who also list White could be compared with other groups.

Survey non-response has the potential to bias survey estimates. Survey response rates are declining and have been declining for some time; this conclusion is valid internationally for all survey types (de Leeuw and de Heer, 2002). During the eleven year period, the TUS-CPS yearly response rates decreased 5% from an initial (1992-1993) rate of 88.0% to a final (2003) rate of 83.0%. Standard adjustments were made in the TUS-CPS statistical weights to correct for potential non-response bias. However, it is still possible that there is some non-response bias in the current smoking prevalence estimates.

Survey non-coverage has the potential to bias survey estimates. The TUS-CPS as well as the NHIS and other general surveys of the U.S. population do not cover all geographic areas; thus, they do not provide full coverage of AIAN as do special studies focusing only on this special population. However, CPS coverage ratios for the residual race category, which includes the NH AIANs, is comparable to those from other race/ethnicities and is more than 80% for both genders (Current Population Survey, 2006 Attachment 16).

Acknowledgements

5. References

Botman SL, Moore TF, Moriarity CL, and Parsons VL (2000). Design and estimation for the National Health Interview Survey, 1995-2004. National Center for Health Statistics. *Vital Health Stat* 2(130).

Bowles M, Ilg RE, Miller S, Robison E, Polivka A. (2003). Revisions to the Current Population Survey effective in January 2003. *Employment and Earnings*.

Caraballo RS, Giovino GA, Pechacek TF, Mowery PD. (2001). Factors associated with discrepancies between self-reports on cigarette smoking and measured serum cotinine levels among persons aged 17 years and older. *Am J Epidemiology*, 153, 807-814.

Centers for Disease Control and Prevention (CDC). Annual smoking-attributable mortality, years of potential life lost, and economic costs - United States, 1995-1999. *Morbidity and Mortality Weekly Report* 51(14):300-303, 2002.

Current Population Survey: Design and Methodology, Technical Paper 63, Issued March 2002 (http://www.census.gov/prod/2002pubs/tp63rv.pdf).

Current Population Survey: February, June, and November 2003: Tobacco Use Supplement File, November 2006 (http://www.census.gov/apsd/techdoc/CPS/cpsmain.html).

Davis WW, Hartman AM, Gibson JT. Bridging estimates by race for the Tobacco Use Supplement to the Current Population Survey. (2007). http://riskfactor.cancer.gov/studies/tus-cps/race_bridging_071307.

de Leeuw, E. and de Heer, W. (2002) Trends in Household Survey Nonresponse: a longitudinal and international comparison. pg 121-134 in *Survey Nonresponse,* (edited by R.M. Groves, D.A.Dillman, J.L. Eltinge, and R.J.A. Little). John Wiley, New York.

Fagan P, Moolchan ET, Lawrence D, Fernander A, Ponder PK (2007a). Identifying health disparities across the tobacco continuum. *Addiction*, 102 (Supp 2), 5-29.

Fagan P, Shavers V, Lawrence D, Gibson JT, and Ponder P. (2007b). Cigarette smoking and quitting behaviors among unemployed adults in the United States. *Nicotine & Tobacco Research*, 9, 241-248.

Fay RE (1984). Some properties of estimates of variance based on replication methods. *American Statistical Association. Proc. Of Sec. Survey Research Methods*, 495-500.

Gelman A and Rubin DB (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-511.

Gilpin EA, Pierce JP, Cavin SW, Berry CC, Evans NJ, Johnson M, Bal DG (1994). Estimates of population smoking prevalence: self- vs proxy reports of smoking status. *Am J Public Health* 84(10), 1576-1579.

Graubard BI, Rao SR, and Gastwirth JL (2005). Using the Peters-Belson method to measure health care disparities from complex survey data. *Statistics in Medicine*, 24, 2659-2668.

Harper S, Lynch J. (2005). *Methods for Measuring Cancer Disparities: Using Data Relevant to Healthy People 2010 Cancer-Related Objectives.* NCI Cancer Surveillance Monograph Series, Number 6. Bethesda, MD: National Cancer Institute, NIH Publication No. 05-5777.

Hyland A, Cummings KM, Lynn WR, Corle D, Giffen CA (1997). Effect of proxy-reported smoking status of population estimates of smoking prevalence. *Am J Epidemiology* 145(8), 746-751.

Korn, EL and Graubard, BI. (1999). *Analysis of Health Surveys*. New York: John Wiley & Sons.

Levy DT, Romano E, Mumford E. (2005). The relationship of smoking cessation to sociodemographic characteristics, smoking intensity, and tobacco control policies. *Nicotine Tob Res* ;7:387-96.

National Cancer Institute (2007). Tobacco Use Supplement: Current Population Survey. (http://riskfactor.cancer.gov/studies/tus-cps).

Neter J and Wasserman W (1974). *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*. Richard D. Irwin, Homewood, Illinois.

Office of Management and Budget. (1997). Revisions to the standards for the classification of Federal data on race and ethnicity. *Federal Register* 62FR58781-58790.

Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, and Kinne S. (1994). The validity of self-reported smoking: a review and meta-analysis. *Am J Public Health*, 84, 1086-1093.

Research Triangle Institute (2004). *SUDAAN Language Manual*, Release 9.0 Research Triangle Park, NC: Research Triangle Institute.

Rubin DB. (1987). *Multiple imputation for nonresponse in Surveys*. John Wiley, New York.

Shavers VL, Lawrence D, Fagan P, and Gibson JT (2005). Racial/ethnic variation in cigarette smoking among the civilian US population by occupation and industry, TUS-CPS 1998-1999. *Preventive Medicine*, 41, 597-606.

Shavers VL, Fagan P, Jouridine Alexander LA, Clayton R, Doucet J, and Baezconde-Garbanati L. (2006). Workplace and home smoking restrictions and racial/ethnic variation in the prevalence and intensity of current cigarette smoking among women by poverty status, TUS-CPS 1998-1999 and 2001-2002. *Epidemiology Community Health* 60 (Supp 2) 44-50.

Tucker C, Miller S, Parker J. (2002). Comparing census race data under the old and new standards. Chapter 19 in *The New Race Question: How the Census counts multiracial individuals.* J. Perlmann and M Waters (editors). Russell Sage Publications.

U.S. Department of Health and Human Services. (1998). *Tobacco Use Among U.S. Racial/Ethnic Minority Groups—African Americans, American Indians, and Alaska Natives, Asian Americans and Pacific Islanders, and Hispanics: A Report of the Surgeon General*. Atlanta, GA: HHS, PHS, CDC, NCCDPHP, OSH.

U.S. Department of Health and Human Services. (2000). *Healthy People 2010: Understanding and Improving Health*. Washington, DC: U.S. Department of Health and Human Services.

U.S. Department of Health and Human Services. (2004). T*he Health Consequences of Smoking: A Report of the Surgeon General.* Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S. Department of Health and Human Services. (2007). Data and Statistics. Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, (http://www.cdc.gov/tobacco/data_statistics).