National Cancer Institute

Tips & Tricks of Analyzing the TUS Data

William W. Davis (NCI)

Anne Hartman (NCI) and Todd Gibson (IMS)

October 23, 2007

U.S. DEPARTMENT
OF HEALTH AND
HUMAN SERVICES

National Institutes
of Health

Talk Outline

1. TUS-CPS variance estimation choices
   - Example using generalized variance function (GVF)
2. Change in CPS race reporting in 2003
   - How is TUS-CPS dealing with the change?
   - Some TUS-CPS trend results and analyses
3. TUS-CPS (Feb 2002 and Feb 2003) overlap sample
   - Derivation and properties of new statistical weights for the overlap sample

Methods of Variance Estimation for TUS-CPS

1. Generalized variance functions (GVF)
   - Fast but only approximate
   - Useful for monthly CPS labor force estimates
2. Balanced repeated replication (BRR) based on replication weights
   - Rep weights not on TUS-CPS public use file (available from NCI on request)
   - Takes time to develop but worth the effort
   - Provides more defensible variance estimate

Variance calculation using GVF

- GVF assumes variance is related to expected value
  - Modeled in terms of parameters, “a” and “b”
  - Parameters are estimated using historical data
- TUS-CPS Source & Accuracy Statement Contains
  - Tables of parameters
  - Examples of variance estimation using GVF
- Standard errors based on GVFs
  - Reasonable for means, totals, percentages and their differences
  - Not available for regression

GVF Example

- Using TUS-CPS 2003, estimate of current smoking percentage, p, for males 18+ = 20.69%.
- Problem: Estimate standard error, s, using GVF:
  - general formulae given for mean, percentage, and total in Source & Accuracy (S&A) document (involve *a* and *b*)

Solution: $s = \left(b * p * (100 - p) / x\right)^{1/2}$

where *x*=total population size
and *b* obtained from S&A lookup table.

$x =$ 101,244,033 (number of males age 18+)

$b =$ 1,575 from Table 5 (S&A table).

GVF example continued (TUS-CPS 2003)

$$s = \left(b * p * (100 - p) / x\right)^{1/2}$$

$$s = \left(\frac{1,575 * 20.69 * (100 - 20.69)}{101,244,033}\right)^{1/2}$$

- Standard error using GVF = 0.160
- Standard error using rep. wgt. = 0.186
- Confidence interval for percentage estimate (11.6%) shorter using GVF

Variance Estimation Summary

- Two methods of variance estimation
  - GVF with public use weights
  - Using replication weights (available on request)
- Showed GVF variance estimation example
  - Compared with estimate using replication weights
- GVF method can be used for smoking prevalence estimation
  - Not as precise as variance using replicate weights

National Cancer Institute

Change in CPS race/ethnicity reporting and the Use of "race/bridging"

U.S. DEPARTMENT
OF HEALTH AND
HUMAN SERVICES

National Institutes
of Health

Change in Standards for reporting race and ethnicity

Office of Management. & Budget (OMB) in 1997 modifies Directive 15, OMB (1977)

1. Federal agencies must report tabulations for

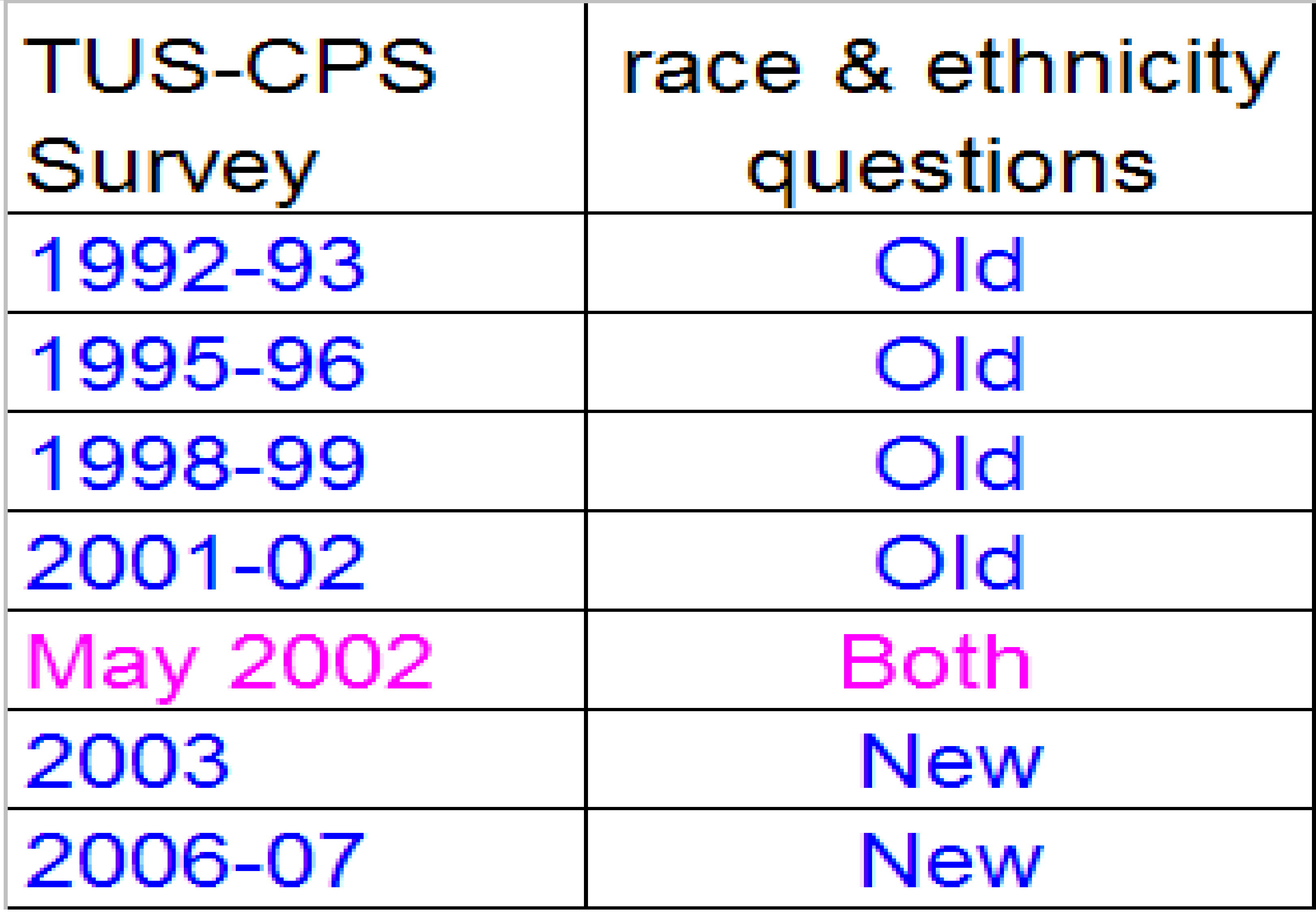
   White, Black, Asian or Pacific Islander (API), American Indian or Alaskan Native (AIAN)

2. Should allow multiple race reporting
3. Hispanic origin should be reported separately
4. Changes should be implemented by 2003

How did TUS-CPS deal with the mandated change in race/ethnicity questions?

1. BLS developed new questions for race/ethnicity
2. BLS sponsored May 2002 CPS Supplement
3. Census tabulated CPS race/ethnicity responses from May 2002 using both “old” and “new” questions
4. NCI used these responses to create a “race bridge”

TUS-CPS race/ethnicity questions



TUS-CPS Survey	race & ethnicity questions
1992-93	Old
1995-96	Old
1998-99	Old
2001-02	Old
May 2002	Both
2003	New
2006-07	New

CPS race/ethnicity questions

Prior to January 2003	Starting in January 2003
What is your race?	Are you Spanish, Hispanic, or Latino?
Respondents are shown a flash card with:	Yes
RACE	No
1. White	
2. Black	Please choose one or more races that you
3. American Indian, Eskimo, or Aleut	consider yourself to be
4. Asian or Pacific Islander	*Respondents are shown flash card with*
What is your origin or descent? [2]	**CHOOSE ONE OR MORE**
Respondents are shown a flash card with:	White
ORIGIN OR DESCENT	Black or African American
10 Mexican-American 14 Puerto Rican	American Indian or Alaska Native
11 Chicano 15 Cuban	Asian
16 Central or South American 17 Other Hispanic	Native Hawaiian or Other Pacific Islander

Major changes to CPS race/ethnicity

- Respondents may now select more than one race when answering the survey.
- Asian or Pacific Islander (API) category split:
    1. Asian
    2. Native Hawaiian or Other Pacific Islander (NHOPI)
- The ethnicity question asked directly whether the respondent was Hispanic.

May 2002 CPS Supplement: “New” Summary

CPS Race/ethnicity	Total	Percent
Hispanic	10,490	12.0%
NH White only	83,877	71.1%
NH Black Only	9,857	11.3%
NH AIAN Only	1,065	0.5%
NH Asian Only	3,712	3.9%
NH NHOPI Only	349	0.2%
NH White-Black	121	0.1%
NH White-Asian	167	0.1%
NH White-AIAN	1,138	0.7%
NH Black-AIAN	130	0.1%
NH Others	269	0.1%
Total	**111,175**	**100.0%**

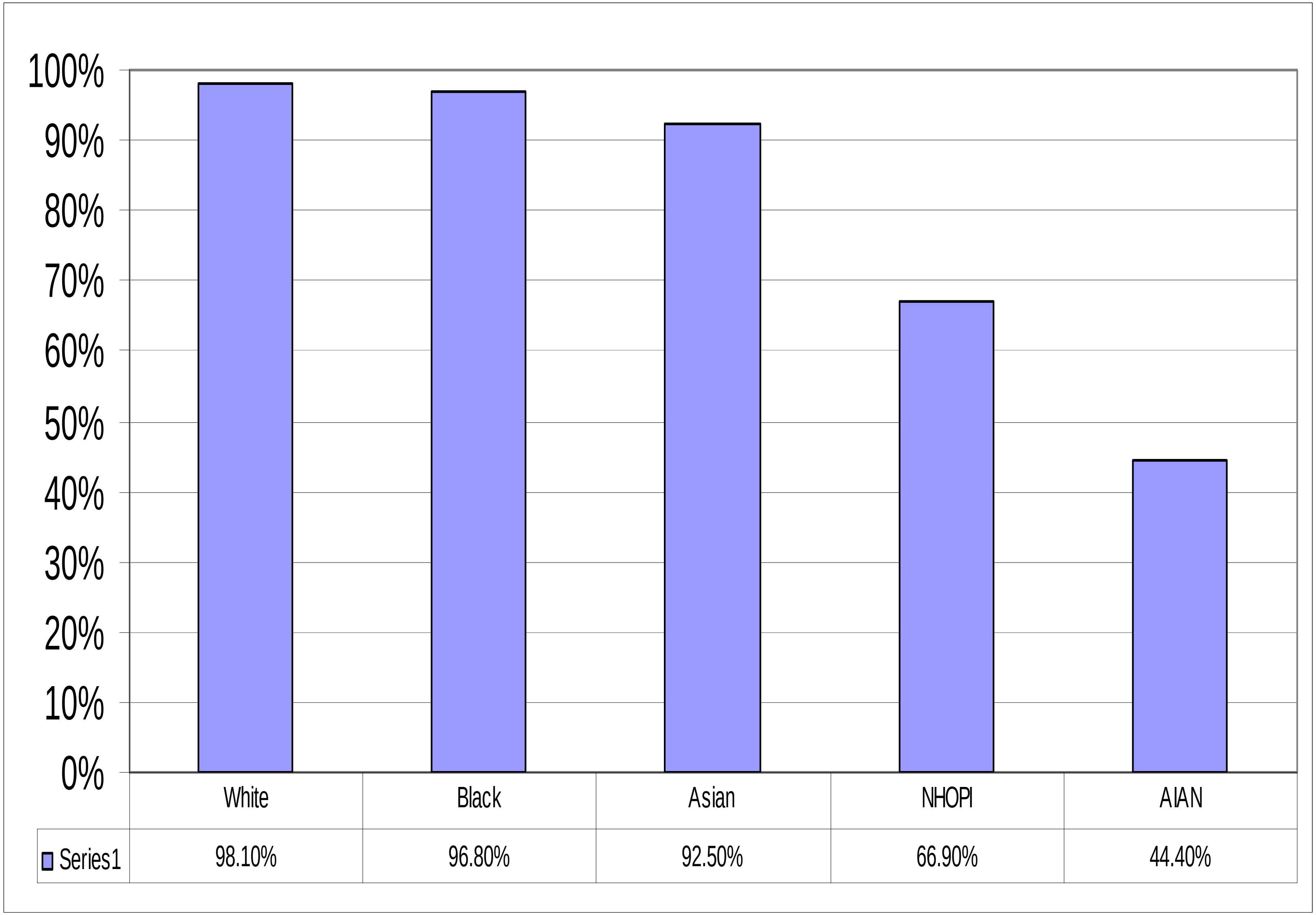
Three estimation methods for TUS-CPS post-2003 race/ethnicity groups

1. Use single race = “only” category
2. Use “any mention” category

❖ Neither of these groups is exactly comparable to pre-2003 group

3. Using May 2002 sample results, develop a model to infer pre-2003 race/ethnicity

Ratio of single race/any mention for non-Hispanics from CPS May 2002

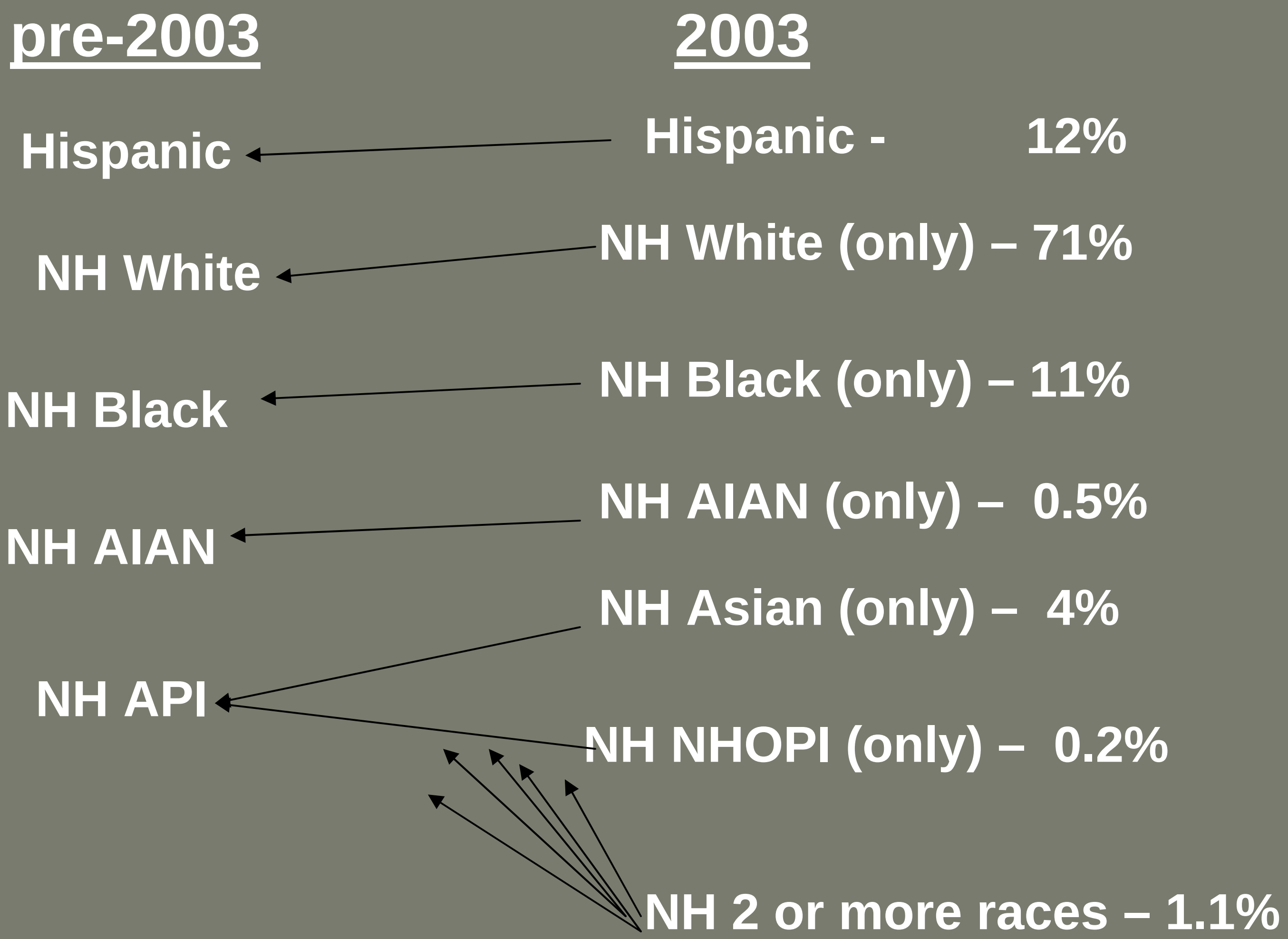


	White	Black	Asian	NHOPI	AIAN
🟪 Series1	98.10%	96.80%	92.50%	66.90%	44.40%

TUS-CPS Race bridging approach

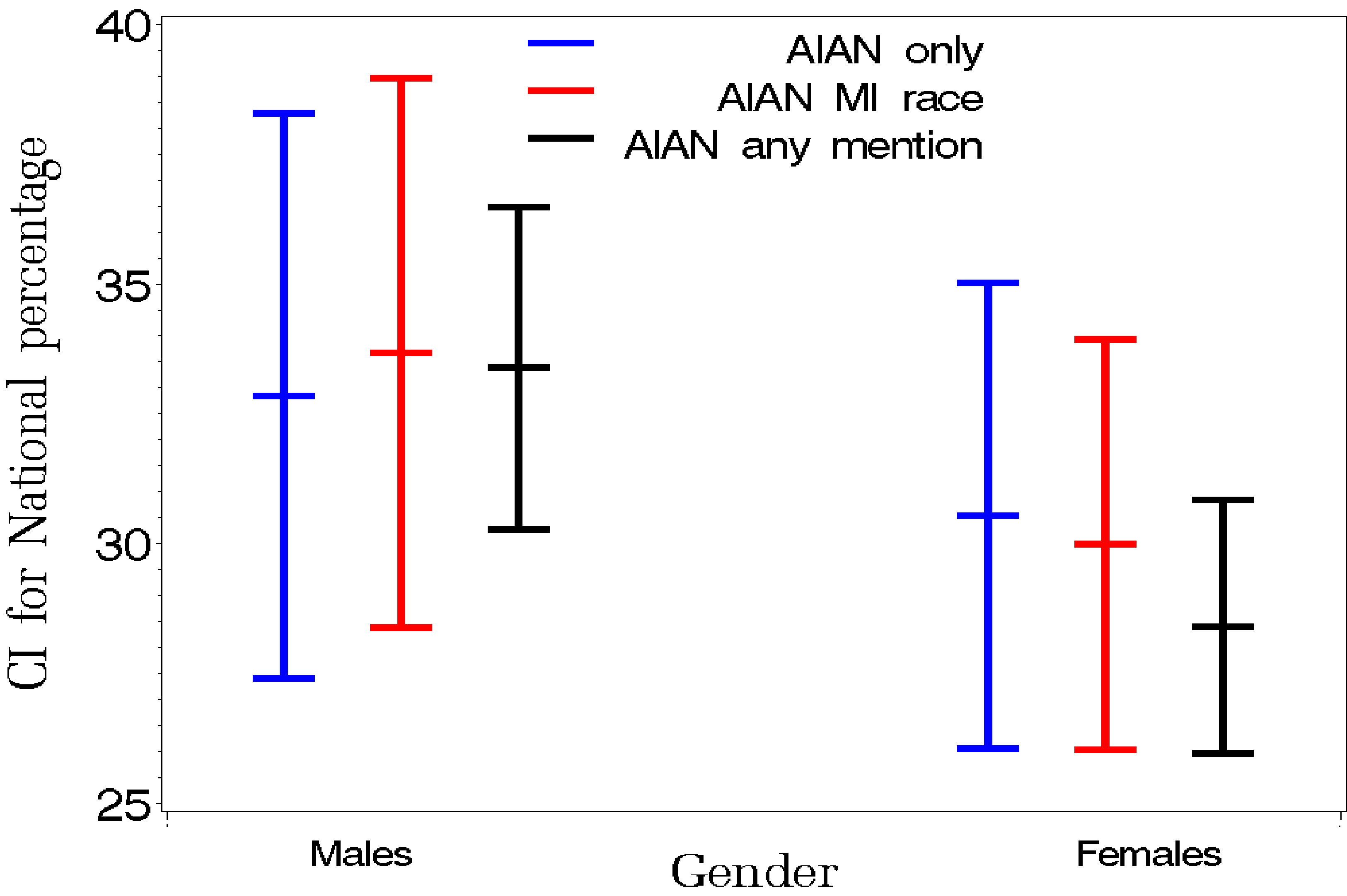
- Simpler version of NCHS approach
  - Schenker and Parker (2003) *Stat. in Med.,* 22, 1571-1587.
- Use May 2002 CPS data (supplied by Census)
  - Develop model to predict pre-2003 race/ethnicities given post-2003 value
- Multiply Impute pre-2003 race/ethnicities for multiple race responders (Rubin, 1987)
- Paper summarizing the approach on website (http:/riskfactor.cancer.gov/studies/tus-cps/race bridging 071307.pdf).

Use of post-2003 race/ethicity to infer pre-2003 race/ethnicity

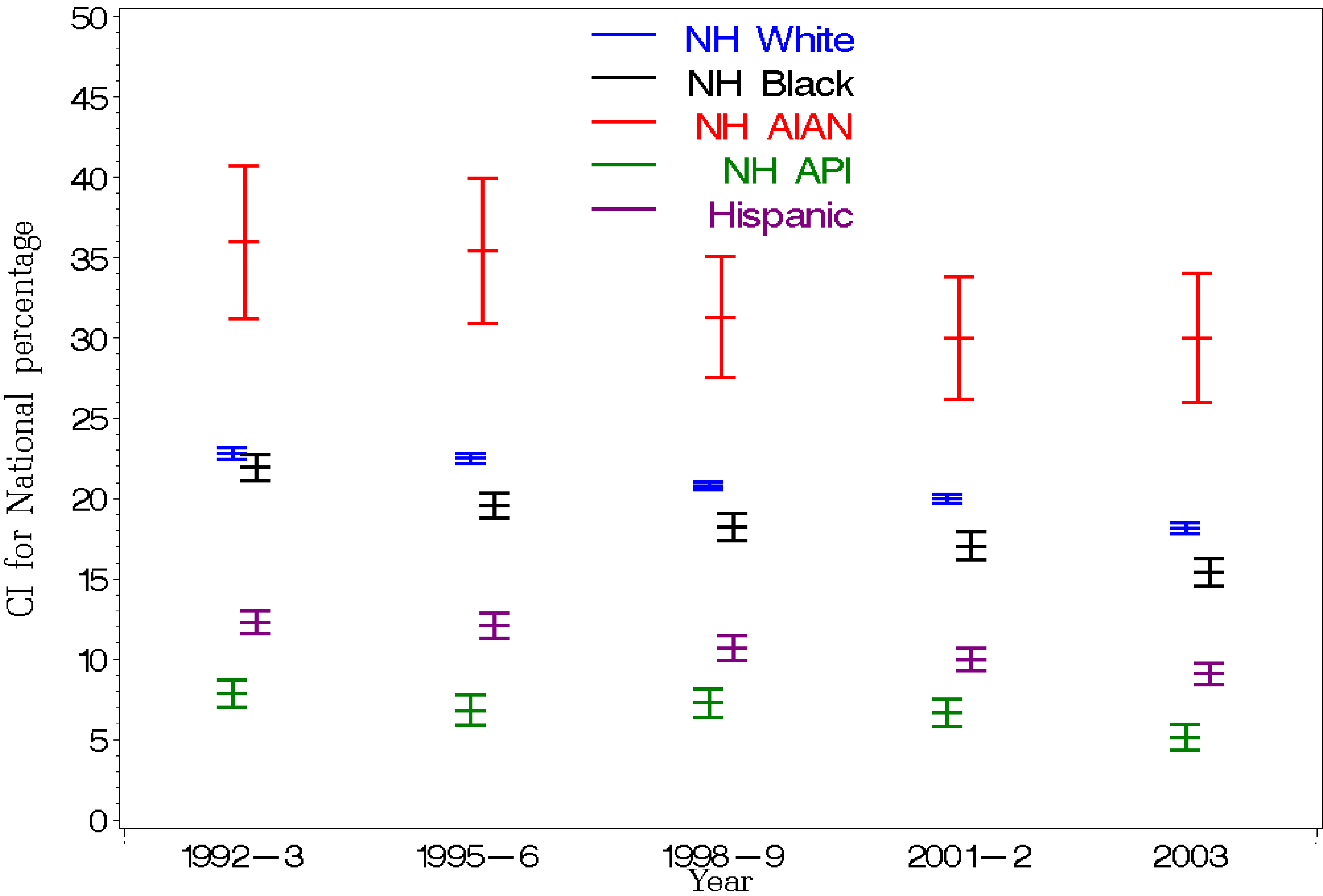


pre-2003	2003
Hispanic	Hispanic - 12%
NH White	NH White (only) – 71%
NH Black	NH Black (only) – 11%
NH AIAN	NH AIAN (only) – 0.5%
NH API	NH Asian (only) – 4%
	NH NHOPI (only) – 0.2%
	NH 2 or more races – 1.1%

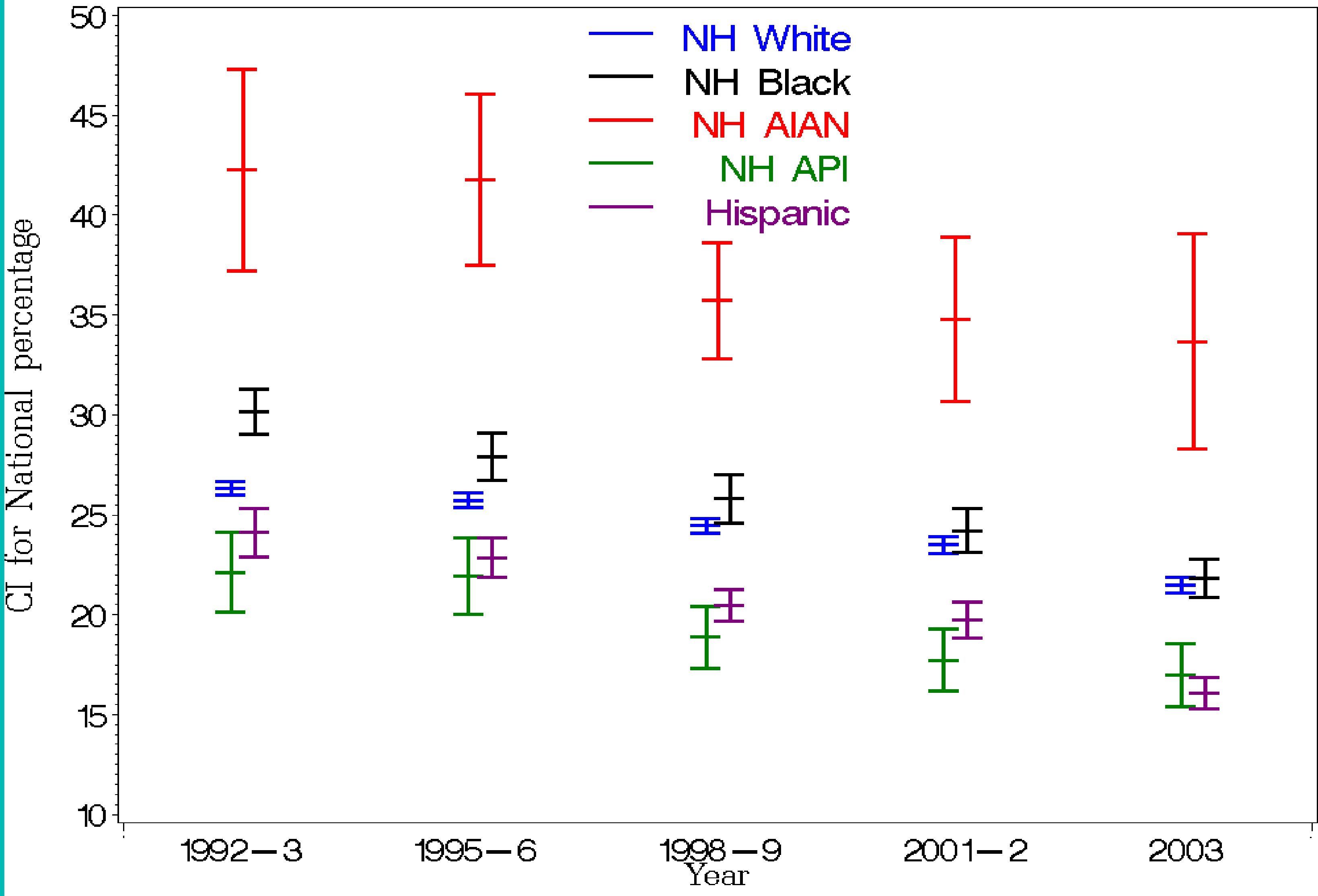
Comparison of three AIAN TUS-CPS current smoking estimates for 2003



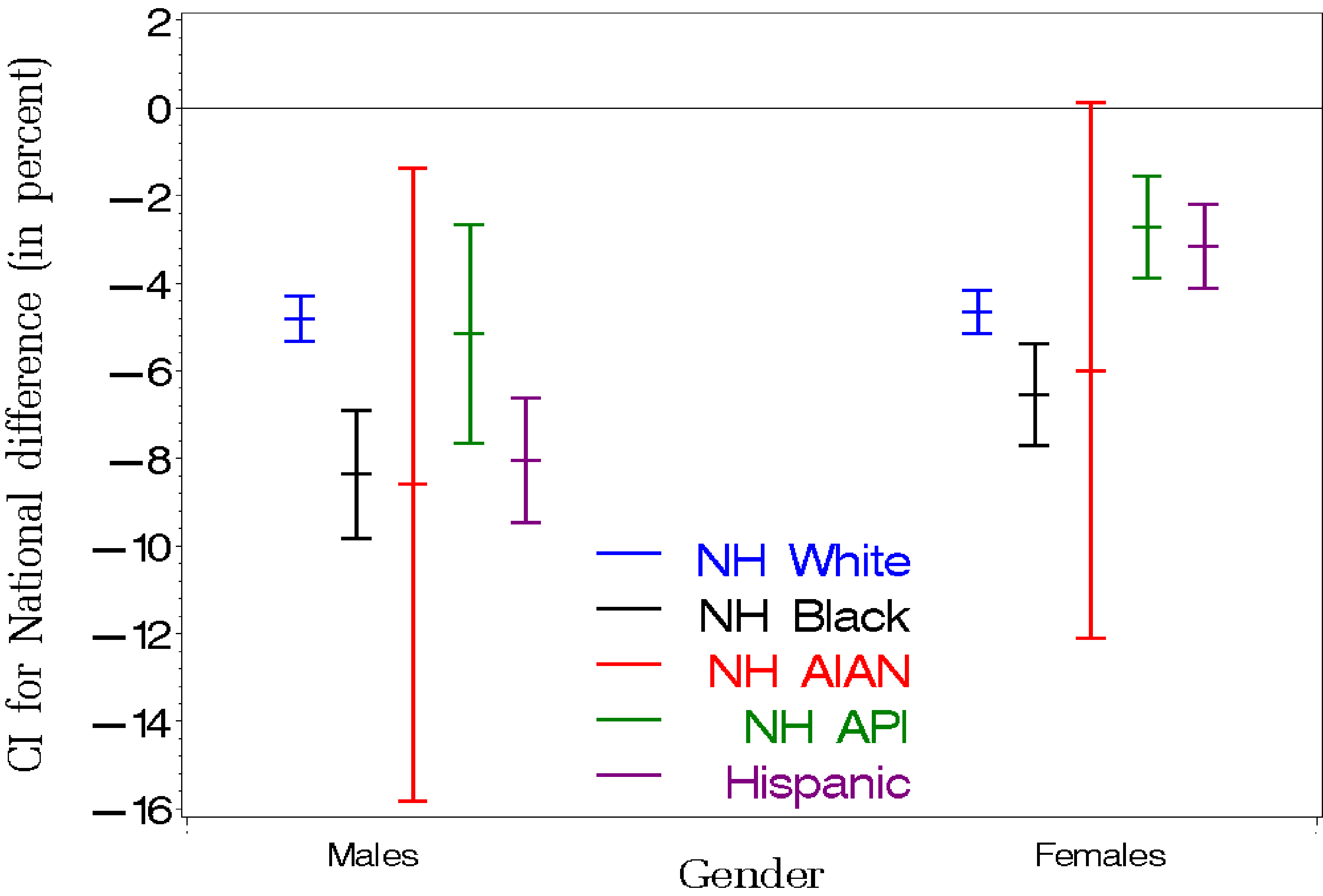
Female Current Smoking Trend from TUS-CPS by race/ethnicity: 1992-2003



Male Current Smoking Trend from TUS-CPS by race/ethnicity: 1992-2003



Difference in Current Smoking by race/ethnicity and gender: 2003 - 1992



Reporting Change for Race/ethnicity Summary

- We described the change in race/ethnicity questions that occurred in 2003
- We develop a “race bridging” technique and apply it to TUS-CPS current smoking
- Most useful for races where a high proportion report multiple races (AIAN)
- We apply to AIAN since they have high current smoking rates

National Cancer Institute

TUS-CPS overlap sample and weighting

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Origin of Overlap Sample

- CPS 4/8/4 panel design
- Persons in sample
  - For TUS-CPS in Feb. 2002 and also
  - For TUSCS-CPS in Feb. 2003
  - 3 Panels satisfied this requirement
- Overlap sample: those who responded to both these surveys
- Responses to the overlap sample can be analyzed as a longitudinal study

Overlap weight modification

- Either the Feb. 2002 or Feb. 2003 stat. weights could be used to construct overlap sample population estimates
  - Either of these analyses would be biased
- To eliminate bias, we adjusted the weights of the overlap sample
  - Adjust for differential non-response by gender, race/ethnicity, age, and geography

Comparison of Counts: Overlap and 2003 TUS

	Unweighted Counts				Weighted Counts		
	Overlap	Feb. 2003	Overlap Percent		Overlap	Feb. 2003	Overlap Percent
All	22,598	68,954	32.8%		71,752,091	224,088,640	32.0%
Hispanic	1,771	6,684	26.5%		7,309,211	27,812,152	26.3%
NH White	17,947	52,152	34.4%		53,784,871	157,866,726	34.1%
NH Black	1,844	6,129	30.1%		7,442,553	25,454,962	29.2%
NH Other	1,036	3,989	26.0%		3,215,456	12,954,800	24.8%

- Differential overlap percent (shown for race/ethnicity above) indicates need for weight adjustment
- Similar differences by age groups also (not shown).

General method to derive overlap sample weights

- Could apply weight adjustment to either Feb 2002 or Feb 2003 statistical weights to obtain overlap sample weight

  $$w^* = r * w$$

  Overlap wgt = (adjustment factor) * (2003 stat wgt)

- Picked Feb 2003 since based on more recent control totals from Census 2000
  - Derived full sample and replicate weights using this method

Derivation of Adjustment factor

- Choose adjustment factor so that sums of overlap sample weights match sums of 2003 sample weights in groups defined by
  - Census region (4)
  - Gender (2)
  - Race/ethnicity (4)
  - Age categories (19)
- Details in http://riskfactor.cancer.gov/studies/tus-cps/TUS-CPS_overlap.pdf.

Variance/bias tradeoff

- Standard variance/bias sampling tradeoff
  - Use of overlap sample weights reduces bias
  - But increases variance
- Estimated increase in length of confidence intervals
  - Non-response replicate weights (11%)
  - Self-response replicate weights (22%)

TUS-CPS Overlap Summary

- TUS-CPS overlap sample of over 22,000 provides a unique tobacco research opportunity
- Overlap sample weights
  - Described the need and the general method of construction
  - The loss from increase in variance seems small in comparison to gain from the bias reduction