

# Tobacco Use Supplement To The Current Population Survey

Users' Workshop

June 2009

Tips and Tricks of Handling the TUS Data

James "Todd" Gibson

Information Management Services, Inc

# Tips and Tricks of Handling the TUS Data

- Basic Information on Getting Started with the Data
- Merging Replicate Weights
- Working with Multiple Years
- Merging Overlap
- Linking to Other CPS Files/Other Supplements to the CPS

# Tips and Tricks of Handling the TUS Data

## Basic Information on Getting Started with the Data

- Public Use Files are ASCII Text Files
- Technical Documentation Included with The Data
  - Overview of the Current Population Survey (CPS)
  - Overview of the Tobacco Use Supplement (TUS)
  - Record Layout of the File
  - TUS Questionnaire
  - Source and Accuracy Statement
- Core and Supplement Variables
  - Core: State and Other Geographic Information, Family Income, Race, Origin, Gender, Age, Education, Marital Status, Labor Force Information, Occupation
  - Supplement: Language of Interview, Interview Method (Telephone/In Person), Relationship of Proxy, Cigarette Smoking Prevalence, Smoking History, Medical and Dental Advice to Quit, Use of Other Forms of Tobacco, Smoking Policy at Home, Workplace Smoking Policy

# Tips and Tricks of Handling the TUS Data

## Basic Information on Getting Started with the Data

- Better Estimates Using All 3 Months of Data Collection
- Supplement Weights
  - Non-Response Weights: Analyses that Included Both Self and Proxy Respondents
  - Self-Response Weights: Analyses of Self Respondents Only
- Example Using May 2006, August 2006 and January 2007 Data
  - Variables: Age, Gender, Race/Ethnicity, Education, Interview, PRS64 (Who is Responding to the Supplement), Self and Non-Response Weights
  - GESTFIPS: State (FIPS)
  - Selections: PRPERTYP=2, PEAGE>17, INTRVIEW=1
  - Since Each Survey is Weighted to the Population, Divide Weights by 3
  - Table of Current Smoking Prevalence Rates by Gender and by Age

Tobacco Use Supplement to The Current Population Survey  
 May 2006, August 2006 and January 2007 Combined  
 Interviewed, Adults, Ages 18 and Over  
**Self and Proxy Respondents**

Current Cigarette Smoking Status By Gender and Age Group

Self and Proxy Respondents	Smoker Recode Never/Former Sample	Smoker Recode Never/Former Population	Smoker Recode Never/Former Percent	Smoker Recode Current Sample	Smoker Recode Current Population	Smoker Recode Current Percent	Total Sample	Total Population	Total Percent
Total	186,723	180,059,805	82.28	40,705	38,765,650	17.72	227,428	218,825,454	100.00
Gender Male	85,366	84,256,721	79.94	21,426	21,142,935	20.06	106,792	105,399,656	100.00
Gender Female	101,357	95,803,084	84.46	19,279	17,622,715	15.54	120,636	113,425,798	100.00
Age Group 18-24	20,714	22,852,427	82.07	4,727	4,992,010	17.93	25,441	27,844,437	100.00
Age Group 25-44	63,764	64,821,609	79.63	17,008	16,580,926	20.37	80,772	81,402,535	100.00
Age Group 45-64	65,021	59,847,003	80.72	15,589	14,291,634	19.28	80,610	74,138,637	100.00
Age Group 65+	37,224	32,538,765	91.81	3,381	2,901,080	8.19	40,605	35,439,845	100.00

Tobacco Use Supplement to The Current Population Survey  
 May 2006, August 2006 and January 2007 Combined  
 Interviewed, Adults, Ages 18 and Over  
**Self Respondents Only**

Current Cigarette Smoking Status By Gender and Age Group

Self Respondents Only	Smoker Recode Never/Former Sample	Smoker Recode Never/Former Population	Smoker Recode Never/Former Percent	Smoker Recode Current Sample	Smoker Recode Current Population	Smoker Recode Current Percent	Total Sample	Total Population	Total Percent
Total	139,827	179,047,732	81.51	31,501	40,605,950	18.49	171,328	219,653,682	100.00
Gender Male	59,295	83,970,198	79.32	15,133	21,897,259	20.68	74,428	105,867,457	100.00
Gender Female	80,532	95,077,534	83.56	16,368	18,708,691	16.44	96,900	113,786,225	100.00
Age Group 18-24	11,158	22,221,636	78.84	3,134	5,965,267	21.16	14,292	28,186,902	100.00
Age Group 25-44	47,690	64,699,376	79.21	13,125	16,979,254	20.79	60,815	81,678,630	100.00
Age Group 45-64	50,019	59,555,978	80.15	12,426	14,745,799	19.85	62,445	74,301,777	100.00
Age Group 65+	30,960	32,570,742	91.78	2,816	2,915,630	8.22	33,776	35,486,373	100.00

# Tips and Tricks of Handling the TUS Data

## Merging Replicate Weights

- Two Replicate Weight Files For Each Survey in 1992-2007
  - Non-Response Weights
  - Self Response Weights
- 48 Replicate Weights in 1992-1993 Files
- 80 Replicate Weights in 1995-2003 Files
- 160 Replicate Weights in 2006-2007 Files
- Unique Identifiers
  - 1992-1993: Household ID (H\_ID) and Persons' Line Number (A\_LINENO)
  - 1995-1996, 1998-1999: Household ID (HRHHID), Serial Suffix (HRSERSUF), Persons' Line Number (PULINENO)
  - 2001-02, 2003, 2006-07: Unique Household Identifier (QSTNUMBER), Unique Person Identifier (OCCURNUM)

# Tips and Tricks of Handling the TUS Data

## Merging Replicate Weights

- Example Merging Replicate Weights and Calculating Current Smoking Prevalence Using SAS and SUDAAN Using the May 2006, August 2006 and January 2007 Data.
  - Read in Main Survey Data
  - Read In Replicate Weight Data. Multiple Lines Per Record.
  - Sort and Merge Each Survey By QSTNUM and OCCURNUM
  - Divide weights by 3
  - Using Replicate Weights in SUDAAN to Calculate Current Smoking Prevalence, Standard Errors and 95% Confidence Intervals
    - Design=BRR (Balance Repeated Replication)
    - ADJFAY=4
  - Table of Current Smoking Prevalence Rates by Gender



**Variance Estimation Method: BRR**  
**Tobacco Use Supplement to The Current Population Survey**  
**May 2006, August 2006 and January 2007**  
**Interviewed, Adults, Ages 18 and Over**  
**Self and Proxy Respondents**  
**Current Cigarette Smoking Status By Gender and Age Group**  
**by: Gender, Current Cigarette Smoker.**

Gender	Current Cigarette Smoker Total	Current Cigarette Smoker Yes	Current Cigarette Smoker No
Total Sample Size	227428	40705	186723
Total Population Size	218825454	38765650	180059805
Total Percent	100.0000	17.7153	82.2847
Total Standard Error	0.0000	0.1130	0.1130
Total Lower 95% CI		17.4933	82.0605
Total Upper 95% CI		17.9395	82.5067
Male Sample Size	106792	21426	85366
Male Population Size	105399656	21142935	84256721
Male Percent	100.0000	20.0598	79.9402
Male Standard Error	0.0000	0.1578	0.1578
Male Lower 95% CI		19.7499	79.6267
Male Upper 95% CI		20.3733	80.2501
Female Sample Size	120636	19279	101357
Female Population Size	113425798	17622715	95803084
Female Percent	100.0000	15.5368	84.4632
Female Standard Error	0.0000	0.1239	0.1239
Female Lower 95% CI		15.2937	84.2170
Female Upper 95% CI		15.7830	84.7063

# Tips and Tricks of Handling the TUS Data

## Working Multiple Years of Data

- Example Merging Replicate Weights and Calculating Current Smoking Prevalence Using SAS and SUDAAN Using Data From 2 Survey Time Periods (2003 & 2006-2007)
  - Read In and Merge Main Survey and Replicate Weights
  - To Combined 2003 and 2006-07 Need to Construct a New Set of 240 Replicate Weights
    - 2003: 80 Replicate Weights
    - 2006-2007: 160 Replicate Weights
    - ADJFay=16
  - Table of Current Smoking Prevalence Rates by Gender

**Variance Estimation Method: BRR**

**Tobacco Use Supplement to The Current Population Survey**

**February 2003, June 2003, November 2003, May 2006, August 2006 and January 2007**

**Interviewed, Adults, Ages 18 and Over**

**Self and Proxy Respondents**

**Current Cigarette Smoking Status By Gender and Age Group**

**by: Gender, Current Cigarette Smoker.**

Gender	Current Cigarette Smoker Total	Current Cigarette Smoker Yes	Current Cigarette Smoker No
Total Sample Size	461702	84536	377166
Total Population Size	214976422	38783864	176192558
Total Percent	100.0000	18.0410	81.9590
Total Standard Error	0.0000	0.0911	0.0911
Total Lower 95% CI		17.8622	81.7788
Total Upper 95% CI		18.2212	82.1378
Male Sample Size	216833	44277	172556
Male Population Size	103321845	21047600	82274244
Male Percent	100.0000	20.3709	79.6291
Male Standard Error	0.0000	0.1216	0.1216
Male Lower 95% CI		20.1324	79.3885
Male Upper 95% CI		20.6115	79.8676
Female Sample Size	244869	40259	204610
Female Population Size	111654577	17736264	93918314
Female Percent	100.0000	15.8849	84.1151
Female Standard Error	0.0000	0.0996	0.0996
Female Lower 95% CI		15.6897	83.9178
Female Upper 95% CI		16.0822	84.3103

# Tips and Tricks of Handling the TUS Data

## Merging Overlap Supplement

- Unique CPS Panel Design Feature
  - Each Household in the Sample is Surveyed for Four Consecutive Months (Panels 1-4) and then for Four Consecutive Months (Panels 5-8) Nine Months Later
  - Persons in Panels 1, 2 or 3 in February 2002 were in Panels 5, 6 or 7 in February 2003

# Tips and Tricks of Handling the TUS Data

## Merging Overlap Supplement

- Matching Variables for February 2002 and February 2003 Overlap
  - Household Identifier (HRHHID)
  - Month in Sample (HRMIS)
    - 1 in Feb 2002 = 5 in Feb 2003
    - 2 in Feb 2002 = 6 in Feb 2003
    - 3 in Feb 2002 = 7 in Feb 2003
    - No TUS Items for Panels 4 and 8 in Both 2002 and 2003.
  - Sample Identifier (HRSAMPLE)
  - Serial Suffix (HRSERSUF)
  - Household Number (HUHHNum)
  - Persons' Line Number (PULINENO)
  - Gender (PESEX)
  - Persons' Age (PEAGE):
    - Match If 2003 Age is within Plus or Minus 1 of 2002 Age

# Tips and Tricks of Handling the TUS Data

## Merging Overlap Supplement

- Match Results
  - 22,598: Self and Proxy
  - 15,846: Self Only
- Reasons For Mismatches
  - Migration: Entire Households and Individuals Move to Another Location
  - Individual or Household Non-Response

# Tips and Tricks of Handling the TUS Data

## Linking to Other CPS Data/Supplements to the CPS

- TUS-CPS Can Be Linked With Other CPS Basic and CPS Supplement Data.
- Allows the Opportunity to Include Other Topics in Analysis Not Included in The TUS
  - March ASEC, Internet Use, Veterans Information, Food Service Information
- Example: Starting in January 2003, Occupation and Industry Were Coded Differently than Pre 2003.
  - Questions Were Not Modified
  - Information Gathered Classified According to New Standards and Definitions
  - How to Group New Occupation Coding Scheme into Occupation Groups White Collar, Blue Collar, Service and Other.
  - Merged February 2002 CPS Data With Bureau of Labor Statistics Monthly Extract File for February 2002.

# Tips and Tricks of Handling the TUS Data

## Merging January 2007 TUS-CPS with 2007 ASEC

- Example Merging the March 2007 Annual Social and Economic (ASEC) Supplement with the January 2007 Tobacco Use Supplement to The Current Population Survey
- Construct Health Insurance Coverage Status (HIC\_Stat) Variable Using ASEC Data. Based on Census Programming Code.
  - <http://www.census.gov/hhes/www/hlthins/hlthinsrecodevar.html>
  - <http://www.census.gov/hhes/www/hlthins/hlthinsvar.html>
- Construct Smoking Ban at Work Variable from TUS-CPS Data
- Tables:
  - Health Insurance Coverage Status, ASEC Data
  - Health Insurance Coverage Status, ASEC/TUS-CPS Merged Data
  - Health Insurance Coverage Status X Smoking Ban at Work, Merged Data



# Tips and Tricks of Handling the TUS Data

## Merging January 2007 TUS-CPS with 2007 ASEC

- Matching Variables for TUS-CPS and ASEC
  - Month in Sample (HRMIS)
    - 1 in January = 3 in March
    - 2 in January = 4 in March
    - 5 in January = 7 in March
    - 6 in January = 8 in March
    - No TUS Items for Panels 4 and 8 in Both 2002 and 2003.
  - PERIDNUM in ASEC (Person Identifier)
  - Construct PERIDNUM in TUS-CPS
    - Household Identifier (HRHHID)
    - Household Identifier 2 (HRHHID2)
    - Persons' Line Number (PULINENO)
  - Gender (PESEX)
  - Persons' Age (PEAGE):
    - Match If January Age is within 1 of March Age

2007 Annual Social and Economic (ASEC) Supplement  
 Adults Ages 18 to 64  
 The FREQ Procedure  
 Health insurance coverage status

HIC_Stat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Covered/Insured	100650	80.89	100650	80.89
Not Covered/Uninsured	23776	19.11	124426	100.00

**January 2007 Tobacco Use Supplement to the Current Population Survey Merged With  
 2007 Annual Social and Economic (ASEC) Supplement**

Adults Ages 18 to 64  
 The FREQ Procedure  
 Health insurance coverage status

HIC_Stat	Frequency	Percent	Cumulative Frequency	<b>Cumulative Percent</b>
Covered/Insured	26864	83.80	26864	83.80
Not Covered/Uninsured	5194	16.20	32058	100.00

**January 2007 Tobacco Use Supplement to the Current Population Survey Merged With  
2007 Annual Social and Economic (ASEC) Supplement  
Adults Ages 18 to 64, Self Respondents**

**The FREQ Procedure**

**Table of HIC\_Stat by WorkBan**

HIC_Stat	Ban	No Ban	Total
Covered/Insured - Frequency	8679	2235	10914
Covered/Insured - Percent	71.06	18.30	89.36
Covered/Insured - Row Pct	79.52	20.48	
Covered/Insured - Col Pct	90.75	84.37	
Not Covered/Uninsured - Frequency	885	414	1299
Not Covered/Uninsured - Percent	7.25	3.39	10.64
Not Covered/Uninsured - Row Pct	68.13	31.87	
Not Covered/Uninsured - Col Pct	9.25	15.63	
Total - Frequency	9564	2649	12213
Total - Percent	78.31	21.69	100.00