



Tobacco Use Supplement

Current Population Survey

CURRENT POPULATION SURVEY, TOBACCO USE SUPPLEMENT TECHNICAL DOCUMENTATION TUS—1992–2015

This file documentation consists of the following:

Section 1 – Overview: Data Harmonization

Section 2 – Overview: Current Population Survey

Section 3 – Overview: Tobacco Use Supplement, 1992–2015

Section 4 – Abstract

Section 5 – Record Layout and Harmonized Variable Listings

[Attachment 1 – Harmonized Data Variable Listing](#)

NOTE: Questions about accompanying documentation should be directed to the National Cancer Institute's Division of Cancer Control and Population Sciences, Behavior Research Program at ncidccpsbrpadvances@mail.nih.gov.

SECTION 1

Overview: Data Harmonization

Data Harmonization

Staff from the NCI TUS team harmonized the data from 1992 through the 2015 waves of the TUS-CPS to provide a more robust data source by increasing the sample size for measures and to allow for more seamless tracking of trends over time. Specifically, the following waves have been harmonized into the 1992-2015 Harmonized Dataset: September 1992, January 1993, and May 1993; September 1995, January 1996, and May 1996; September 1998, January 1999, and May 1999; January 2000 and May 2000; February, June, and November 2003; June 2001, November 2001, and February 2002; May 2006, August 2006, and January 2007; May 2010, August 2010, January 2011, and a matched longitudinal cohort file in May 2010-May 2011; and July 2014, January 2015, and May 2015.

Items that were used in two or more questionnaires were considered eligible for harmonization. The harmonization process involved reviewing each eligible variable across the available survey waves and assigning it into one of the following categories:

- Variables to harmonize as is—variables that remained consistent over time were harmonized in their current state.
- Variables to harmonize after adjustments—items that had slight wording changes over time or those for which the universe had slight variations across survey waves were first adjusted to make the resultant data more complete. All adjustments were documented and can be found in the notes section.
- Variables to drop—items that had significant variations in wording, structure, or universe were not harmonized.

Listings of the variable names, the above categorizations, and decisions made per variable are available in Attachment 1.

For items that were harmonized, the TUS team created one variable name across all waves, using a “flag” variable to keep track of the survey year. The harmonized data include only self-reported responses instead of both self and proxy responses, as most researchers have not typically used proxy responses since the data are very limited and may raise validity issues in some circumstances. The harmonized data file only includes respondents age 18 and older.

Replicate weights for calculating variance estimates will be released in the near future. SAS code and information on linking with the main harmonized data file will be supplied when available.

Harmonized Topics

Core variables from CPS (demographic, geographic, occupational, economic variables) were included in the harmonization. Harmonized topics include:

- Cigarette use, both current and prior usage among daily and nondaily smokers. Also includes menthol use since 2003

- Workplace and home smoking restrictions
- Attitudes toward smoke-free policies in indoor work areas/public places (no harmonization for smoke-free attitudes for multi-unit housing which started in 2014-15 and will later be harmonized with future data)
- Advice by physician/dentist to quit smoking
- Health perceptions/beliefs (harm reduction)
- Smoking history, cessation, former smokers
- Other tobacco products: cigar, pipe, and smokeless tobacco use; other emerging products, including e-cigarettes

The citation for the harmonized data is:

National Cancer Institute. 2019. Tobacco Use Supplement to the Current Population Survey Harmonized Data, 1992-2015. <http://cancercontrol.cancer.gov/brp/tcrb/tus-cps/>.

For more information about past, current, and future NCI TUS, and data reports and publications utilizing supplement data see the NCI website: <http://cancercontrol.cancer.gov/brp/tcrb/tus-cps/>.

Data Collection

Respondents are included in a TUS-CPS sample (wave) typically only once. The rotation of households in and out of sample for the TUS-CPS was generally sufficiently spaced, with months between interviewing so the sample did not contain many overlapping panels. This spacing eliminated most individuals being in the sample twice within the various TUS waves main data files. Exceptions for those cohorts include May 2010-2011, February 2002-2003, January and May 1999, and January and May 2000.

Additional information on data collection can be found in the technical documentation for each individual wave.

SECTION 2

Overview: Current Population Survey

Introduction

The CPS is the source of the official government statistics on employment and unemployment. The CPS has been conducted monthly for over 50 years. Currently, we obtain interviews from about 54,000 households monthly, scientifically selected on the basis of area of residence to represent the nation as a whole, individual states, and other specified areas. Each household is interviewed monthly for four consecutive months one year, and again for the corresponding time period a year later. This technique enables us to obtain reliable month-to-month and year-to-year comparisons at a reasonable cost while minimizing the inconvenience to any one household.

Although the main purpose of the survey is to collect information on employment, a very important secondary purpose is to collect information on demographic characteristics such as age, sex, race, marital status, educational attainment, family relationship, occupation, and industry. From time to time, additional questions are included on health, education, income, and previous work experience. The statistics resulting from these questions serve to update similar information collected every 10 years through the decennial census, and are used by government policymakers and legislators as important indicators of our nation's economic situation and for planning and evaluating many government programs.

The CPS provides current estimates of the economic status and activities of the population of the United States. Because it is not possible to develop one or two overall figures (such as the number of unemployed) that would adequately describe the whole complex of labor market phenomena, the CPS is designed to provide a large amount of detailed and supplementary data. Such data are made available to meet a wide variety of needs on the part of users of labor market information.

Thus, the CPS is the only source of monthly estimates of total employment (both farm and nonfarm); nonfarm self-employed people, domestic workers, and unpaid helpers in nonfarm family enterprises; wage and salaried employees; and, finally, estimates of total unemployment.

It provides the only available distribution of workers by the number of hours worked (as distinguished from aggregate or average hours for an industry), permitting separate analyses of part-time workers, workers on overtime, etc. The survey is also the only comprehensive current source of information on the occupation of workers and the industries in which they work. Information is available from the survey not only for persons currently in the labor force but also for those who are outside the labor force. The characteristics of such persons, whether married women with or without young children, disabled persons, students, older retired workers, etc., can be determined. Information on their current desire for work, their past work experience, and their intentions as to job seeking are also available.

For a more detailed discussion about the basic labor force data gathered monthly in the CPS survey, see "Explanatory Notes and Estimates of Error" in any recent issue of the

Employment and Earnings, a Bureau of Labor Statistics periodical. This source is referred to on the next page.

CPS Sample Design

The CPS is a monthly survey designed primarily to produce national and state estimates of labor force characteristics of the civilian noninstitutional population (CNP) 16 years of age and older. It is conducted in approximately 60,000 eligible housing units throughout the United States. (Note: "Eligible" can be simplistically defined as an occupied housing unit having at least one person in the CNP.) This sample includes 10,000 eligible housing units from the monthly supplementary sample to improve state-level estimates of health insurance coverage for low-income children, also known as the CHIP expansion. This supplementary sample has been part of the official CPS since July 2001. Thirty-two states plus the District of Columbia contain this supplementary sample each month.

The CPS sample is based on information from the 2010 Decennial Census, in accordance with usual practice. Historically, the CPS sample has been redesigned after each Decennial Census.

The CPS sample is a probability sample based on a stratified two-stage sampling scheme: selection of sample primary sampling units (PSUs) and selection of sample housing units within those PSUs. In general, the CPS sample is selected from lists of addresses obtained from the Master Address File (MAF) with updates from the United States Postal Service (USPS) twice a year. The MAF is the Census Bureau's permanent list of addresses, including their geographic locations, for individual living quarters. It is continuously maintained through partnerships with the USPS; federal, state, regional, and local agencies; and the private sector. Also, it is used as a sample frame by many Census Bureau demographic surveys.

Approximately 72,000 housing units are assigned for interview each month, of which about 60,000 are occupied and thus eligible for interview. The remainder are units found to be destroyed, vacant, converted to nonresidential use, containing persons whose usual place of residence is elsewhere, or ineligible for other reasons. Of the 60,000 occupied housing units, approximately 10 percent are not interviewed in a given month due to temporary absence (vacation, etc.), residents not found at home after repeated attempts, inability of persons contacted to respond, unavailability for other reasons, and refusals to cooperate. The interviewed households contain approximately 108,000 people 15 years old and over, approximately 27,000 children 0–14 years old, and about 450 Armed Forces members living with civilians either on or off base within these households. A more precise explanation regarding the CPS sample design is provided in "Explanatory Notes and Estimates of Error: Household Data - Sampling" in any issue of *Employment and Earnings*.

Relationship of Current Population Survey Files to Publications

Each month, a significant amount of information about the labor force is published by the Bureau of Labor Statistics in the *Employment and Earnings* and *Monthly Labor Review* reports.

As mentioned previously, the CPS also serves as a vehicle for supplemental inquiries on subjects other than employment, which are periodically added to the questionnaire. From the basic and supplemental data, the Bureau of the Census issues three series of publications under the general title Current Population Reports:

- P-20 Population Characteristics
- P-23 Special Studies
- P-60 Consumer Income

All Current Population Reports, including the other series for population estimates and projections and special censuses, may be obtained by subscription from the U.S. Government Printing Office at 202-783-3238.

Subscriptions are available as follows: Population Characteristics, Special Studies, and Consumer Income series (P-20, P-23, P-60) combined, \$101 per year (sold as a package only); Population Estimates and Projections, (P-25), \$27 per year. Single issues may be ordered separately; ordering information and prices are provided in the *U.S. Census Bureau Catalog and Guide*, the *Monthly Product Announcement*, and *Census and You*. Selected reports also may be accessed at <http://census.gov/library/publications.html>.

Geographic Limitations

The CPS sample was selected so that specific reliability criteria were met nationally, for each of the 50 states and for the District of Columbia. Since 1985, these reliability criteria have been maintained through periodic additions and deletions in the state samples. Estimates formed for geographic areas identified on the microdata file that are smaller than states are not as reliable.

Weights

Under the estimating methods used in the CPS, all of the results for a given month become available simultaneously and are based on returns for the entire panel of respondents. The CPS estimation procedure involves weighting the data from each sample person. The base weight, which is the inverse of the probability of the person being in the sample, is a rough measure of the number of actual people the sample person represents. Almost all sample people in the same state have the same base weight, but the weights across states are different. Selection probabilities may also differ for some sample areas due to field subsampling, which is done when areas selected for the sample contain many more households than expected. The base weights are then adjusted for noninterview, and the ratio estimation procedure is applied.

1. **Noninterview adjustment.** The weights for all interviewed households are adjusted to the extent needed to account for occupied sample households for which no information was obtained because of absence, impassable roads, refusals, or unavailability of the respondent for other reasons. This noninterview adjustment is made separately for clusters of similar sample areas that are usually, but not necessarily, contained within a state. Similarity of sample areas is based on Core-Based Statistical Area (CBSA) status and size. Within each

cluster, there is a further breakdown by residence. Each CBSA cluster is split by "principal city" and "balance of the CBSA." The proportion of occupied sample households not interviewed fluctuates around 8 percent depending on weather, vacations, etc.

2. **Ratio estimates.** The distribution of the population selected for the sample may differ somewhat, by chance, from that of the population as a whole in such characteristics as age, race, sex, and state of residence. Because these characteristics are closely correlated with labor force participation and other principal measurements made from the sample, the survey estimates can be substantially improved when weighted appropriately by the known distribution of these population characteristics. This is accomplished through two stages of ratio adjustment as follows:
 - a. **First-stage ratio estimate.** The purpose of the first-stage ratio adjustment is to reduce the contribution to variance that results from selecting a sample of PSUs rather than drawing sample households from every PSU in the nation. This adjustment is made to the CPS weights in two race cells: black and nonblack; it is applied only to PSUs that are non-self-representing and for those states that have a substantial number of black households. The procedure corrects for differences that existed in each state cell at the time of the 2000 census between (1) the race distribution of the population in sample PSUs and (2) the race distribution of all PSUs (both 1 and 2 exclude self-representing PSUs).
 - b. **Second-stage ratio estimate.** This procedure substantially reduces the variability of estimates and corrects, to some extent, for CPS undercoverage. The CPS sample weights are adjusted to ensure that sample-based estimates of population match independent population controls. Three sets of controls are used:
 - 1) Fifty-one state controls of the civilian noninstitutional population 16 years of age
 - 2) National civilian noninstitutional population controls for 14 Hispanic and 5 non-Hispanic age-sex categories
 - 3) National civilian noninstitutional population controls for 66 white, 42 black, and 10 "other" age-sex categories

The independent population controls are prepared by projecting forward the resident population as enumerated on April 1, 2000. The projections are derived by updating demographic census data with information from a variety of other data sources that account for births, deaths, and net migration. Estimated numbers of resident Armed Forces personnel and institutionalized people reduce the resident population to the civilian noninstitutional population. Estimates of net census undercount, determined from the Post Enumeration Survey, are added to the population projections. Prior to January 2003, the projections were based on earlier censuses, and prior to January 1994, there was no correction for census undercount. A summary of the current procedures used to make population projections is given in "Revisions in the Current Population Survey Effective January 2003" in the January 2003 issue of *Employment and Earnings*.

SECTION 3

Overview: Tobacco Use Supplement, 1992–2015

General

NCI and the FDA currently co-sponsor the supplement. NCI has sponsored the TUS-CPS regularly since 1992 (more information on individual waves can be found on the NCI website <http://cancercontrol.cancer.gov/brp/tcrb/tus-cps/>). The Centers for Disease Control and Prevention (CDC) was a co-sponsor with NCI from 2001-02 through 2006-07. The CPS is a monthly labor force survey conducted in more than 50,000 interviewed households across the country.

Specifically, to date, the TUS was fielded in 1992-1993, 1995-1996, 1998-1999, 2000, 2001-2002, 2003, 2006-2007, 2010-2011, and 2014-2015. The 2018-2019 cycle is currently being fielded by the Census Bureau. NCI additionally sponsored a special longitudinal TUS-CPS cohort, administered in both May 2010 and May 2011, in addition to the 2010-2011 cycle. Note that part of the 2001-2002 wave (panels 1-3 of February 2002) overlap with part of the 2003 wave (panels 5-7 of February 2003), providing another opportunity to analyze longitudinal data. Likewise, there is overlap between January and May 1999 and January and May 2000, respectively.

TUS data can be used by researchers to monitor tobacco control progress; conduct tobacco-related research, including tobacco health disparities; and evaluate tobacco control programs. Although the TUS has changed slightly between 1992 and 2015, it has generally contained the same information covering:

- current cigarette smoking status and amount smoked;
- use of menthol cigarettes (since 2003);
- smoking history, quit attempts, and intention to quit;
- level of nicotine dependence (since 2003);
- cost of cigarettes and purchase location (since 2003);
- medical/dental advice to quit;
- cigar, pipe, and smokeless tobacco use;
- harm reduction and other emerging products (since 2003);
- workplace and home smoking restrictions; and
- attitudes toward smoke-free policies in public places.

More recent series have included more detailed information on:

- emerging tobacco products;
- use of flavored non-cigarette tobacco products; and
- attitudes toward smoking in multi-unit housing.

SECTION 4

Abstract

The National Cancer Institute (NCI) of the National Institutes of Health (NIH), the principal sponsor of the Tobacco Use Supplement (TUS) series (1992–present) to the Census Bureau’s and the Bureau of Labor Statistics’ Current Population Survey (CPS) has harmonized the TUS and core demographic CPS data to facilitate trend analyses. Some past TUS survey waves have also been co-sponsored with the Centers for Disease Control and Prevention (CDC) and the U.S. Food and Drug Administration (FDA). The basic CPS and its accompanying TUS are conducted by the Census Bureau. Washington: Census Bureau [producer and distributor of the original data format], 1992–2015.

Type of File

Microdata; unit of observation is person.

Universe Description

The universe is person level for all persons aged 18 and older (prior to 2007 also included those 15–17 years old) in the civilian noninstitutional population of the United States who have completed the labor force interview.

Questionnaire Structure

Over the period of 1992–2015, the TUS-CPS has generally been structured in Sections A through K (except I)

- **Section A** screened for prior cigarette usage and current usage status of every-day smoker, some-days smoker, or former smoker (now “not at all” smoking).
- **Section B** questions were asked if the person had smoked 100 cigarettes and was a self-respondent currently smoking every day.
- **Section C** questions were asked if the person had smoked 100 cigarettes and was a self-respondent currently smoking some days.
- **Section D** asked self-respondent current every-day and some-day smokers questions about quit smoking attempts of the past 12 months, and if none, asked if ever quit.
- **Section E** asked self-respondent current every-day and some-day smokers about a varying subset of methods used (including quit lines, Internet or Web-based tools, switching to other tobacco products, nicotine replacement therapy, prescription antidepressants used for cessation treatment, counseling, classes and support groups) during the most recent quit attempt within the past 12 months.
- **Section F** asked self-respondent current every-day and some-day smokers about medical (and often also dental) doctor advice to stop smoking.

- **Section G** asked self-respondent current every-day and some-day smokers about the person's interest, likelihood, and confidence in quitting within the next 6 months.
- **Section H** questions were asked of self-respondent former smokers (if the person had smoked 100 cigarettes, was a self-respondent and currently not smoking at all) similar questions asked of current smokers in Sections B, C, E, and F.
- **Section J** asked all respondents about use of other tobacco products.
- **Section K** asked all self-respondents about smoke-free policies at work and at home, and about attitudes toward smoking in different places.

Subject-Matter Description

Data are provided on labor force activity for the week prior to the survey. Comprehensive data are available on the employment status, occupation, and industry of people 15 years old and older. Also shown are personal characteristics such as age, sex, race, Hispanic origin, marital status, veteran status, immigration status, household relationship, educational background, and some daily living functional status items.

Geographic Coverage

Geography data is provided to the state level and some sub-state levels for specific metropolitan identifiers (see Technical Documentation from individual waves for more details).

Reference Materials

Documentation for individual survey waves is located at <https://cancercontrol.cancer.gov/brp/tcrb/tus-cps/results.html>. Additional information can be found at <https://www.census.gov/programs-surveys/cps/technical-documentation/complete.html>—Census Bureau. *The Current Population Survey Design and Methodology* (Technical Paper 66) describes in detail the sample design and survey procedures used as well as accuracy of estimates and sampling errors. Reference copies should be available from most public libraries or Federal Depository Libraries.

File Availability

The main data files, technical documentation, and replicate weight files are available for download at <https://cancercontrol.cancer.gov/brp/tcrb/tus-cps/results.html>.

SECTION 5

Record Layout and Harmonized Variable Listings

As discussed in Section 1, the harmonized data file contains variables with similar content over the 1992-2015 waves. Data for harmonized variables only includes self-respondents; no data from proxy respondents were included due to potential validity issues. Reasons for variables not being harmonized included not having two or more waves of data available, too many inconsistencies in question wording or structure over time, substantial changes in the universe of respondents, or that a harmonized variable may not be useful to researchers for a specific item (e.g., cigarette brand). The notes column provides details regarding whether a variable was constructed using two or more variables in the single waves (i.e., it is a new variable not found in single wave data), on the universe of respondents included, or if wording or response categories changed over time but not sufficiently to change the meaning/construct of the measure.

Wave Indicator: In the harmonized data file, the variable SURWAVE was constructed as a numerical variable to subset analyses to specific survey waves or to use a flag variable to track estimates over time.

Additional materials and information for users can be found on the NCI website at <http://cancercontrol.cancer.gov/brp/tcrb/tus-cps/>:

- A data dictionary of all harmonized variables and their response categories.
- Details on the contents, number of variables, number of observations, and version of the harmonized file.
- Descriptive tables for all variables in the harmonized file.

SAS programs are provided on the TUS-CPS website to read in the harmonized data and apply variable formats. Additionally, sample code in SAS and SUDAAN are provided below for generating descriptive tables:

SAS

```
Proc surveyfreq data= analysis varmethod=BRR;  
Tables var1 var2 var3 / cl (type=Wilson);  
Weight SRWEIGHT;  
Title = 'PROC SURVEYFREQ';  
Run;
```

```
Proc surveymeans data= analysis varmethod=BRR;  
VAR var4;  
Weight SRWEIGHT;  
Title = 'PROC SURVEYMEANS';  
Run;
```

```
Proc surveylogistic data= analysis varmethod=BRR;  
Model response1 (event='1') = cov1 cov2;
```

```
Weight SRWEIGHT;  
Title = 'PROC SURVEYLOGISTIC';  
Run;
```

```
Proc surveyreg data= analysis varmethod=BRR;  
Model response2 = cov1 cov2/ solution;  
Weight SRWEIGHT;  
Title = 'PROC SURVEYREG';  
Run;
```

For domain analyses add “by domainvar” above the weight statement to subset analyses by your domain variable.

SUDAAN

```
Proc crosstab data=analysis filetype=sas design=wr;  
Weight SRWEIGHT;  
Tables var1 var2 var3;  
Class var1 var2 var3;  
Print / style = nchs tablecell=all;  
Title = 'SUDAAN proc crosstab';  
Run;
```

```
Proc descript data=analysis filetype=sas design=wr;  
Weight SRWEIGHT;  
Var var5;  
Print / style = nchs;  
Title = 'SUDAAN proc descript';  
Run;
```

```
Proc regress data=analysis filetype=sas design=wr;  
Weight SRWEIGHT;  
Model response1 = cov1 cov2;  
Title = 'SUDAAN proc descript';  
Run;
```

```
Proc rlogist data=analysis filetype=sas design=wr;  
Weight SRWEIGHT;  
Model response2 = cov1 cov2;  
Title = 'SUDAAN proc rlogist';  
Run;
```

For domain analyses add “subpopn domainvar = level” above the weight statement to subset analyses by the level of your domain variable.

```
Proc Tabulate Data=Analysis Missing;  
Class SURWAVE <Var1>;  
Var SRWEIGHT;  
Table <Var1> All="Total"  
SURWAVE*SRWEIGHT=""*PctSum<Var1 All>="Percent"*F=7.2;
```

Run;