ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

ICPSR 36498

Population Assessment of Tobacco and Health (PATH) Study [United States] Public-Use Files

United States Department of Health and Human Services. National Institutes of Health. National Institute on Drug Abuse

United States Department of Health and Human Services. Food and Drug Administration. Center for Tobacco Products

ICPSR Public-Use Files User Guide

Inter-university Consortium for Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Terms of Use

Information about Copyrighted Content

NOTE

This document does not currently comply with 508 standards. If you need an alternative means of access to any information, please contact NAHDAP at **nahdap@icpsr.umich.edu**. Let us know the nature of your accessibility problem, the Web address of the required information, and your contact information.

PATH
Population Assessment of Tobacco and Health
A collaboration between the NIH and FDA

Contract #: HHSN271201100027C, HHSN271201600001C

PATH Study Public Use Files
User Guide

Date: August 13, 2019, Updated for Wave 4

Prepared by:
Westat
An Employee-Owned Research Corporation
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

This page is intentionally blank.

Table of Contents

Contents (continued)

Contents (continued)

Contents (continued)

Tables (continued) **Page**

Contents (continued)

Tables (continued) **Page**

Figures

1. Introduction

This User Guide describes the data files included in the Public Use Files (PUFs) for the Population Assessment of Tobacco and Health (PATH) Study. It includes information about the sample design, data collection methods, response rates, and weighting and imputation methods; it also provides guidance for data users on the variance estimation of complex survey data. The PUFs contain extensive data from the adult interview, youth, and parent interviews. Every effort is made to protect the identity of individual respondents. The process of preparing the files for release includes a formal disclosure risk analysis. Any information from the data files that might be used to directly identify respondents is removed and some edits have been made for purposes of confidentiality. Additionally, for a small percentage of cases, values have been swapped across cases with similar characteristics to make it very difficult to identify a respondent with certainty. The modifications used to reduce the likelihood that any respondent could be identified in the data do not affect the overall data quality. However, they may result in estimates that differ slightly from those created using the PATH Study's Restricted Use Files.

In this document, the term "participants" is used to indicate people who agreed to be part of the PATH Study whether or not they completed an interview in a particular wave. The term "respondents" is used to indicate the subset of people who actually completed an interview or, in the case of shadow youth, verified their information with the study in a particular wave; where needed for clarity, this term may be qualified to indicate the type of respondent (e.g., "shadow youth respondent," "adult interview respondent," etc.).

1.1 Background of the PATH Study

On June 22, 2009, the Family Smoking Prevention and Tobacco Control Act (TCA) was signed into law. The TCA gave the Food and Drug Administration's (FDA's) Center for Tobacco Products (CTP) regulatory authority over the manufacture, marketing, and distribution of tobacco products with the goal of protecting the nation's health. This authority allows the CTP to:

- Develop product standards, e.g., regulate levels of nicotine and other ingredients in tobacco products;
- Restrict tobacco marketing and sales to youth;

- Mandate labeling of tobacco products including health warnings on packaging and in advertisements;
- Mandate warning labeling on smokeless tobacco products;
- Ensure "modified risk" claims are supported by scientific evidence;
- Require disclosure of ingredients in tobacco products; and
- Preserve state, local, and tribal authority over tobacco regulation.

In contrast to other Centers within the FDA that review therapeutic products using a "safe and effective" standard, the CTP regulates tobacco products using a standard based on population health. This standard allows CTP to weigh the potential benefits and harm to current, former, and never users of tobacco products when instituting regulations.

The PATH Study is a nationally representative, longitudinal cohort study of tobacco use and how it affects the health of people in the United States. The PATH Study is the first large joint research effort on this topic by the National Institutes of Health (NIH) and FDA since the 2009 TCA authorized the FDA to regulate tobacco products. The PATH Study will provide an empirical evidence base for developing, implementing, and evaluating regulations governing tobacco products by measuring the behavioral and health effects associated with changes in such regulations. The PATH Study interviews tobacco users and nonusers about use of multiple tobacco products, including cigarettes, e-cigarettes, traditional cigars, cigarillos, filtered cigars, smokeless tobacco, snus pouches, pipe tobacco, hookah, dissolvable tobacco and for youth, bidis and kreteks. Biomarkers of exposure and potential harm are measured in the blood and urine specimens of a subsample of participants; these biomarker data can be analyzed with associated data on any tobacco product use as well as product use for the 3 days prior to biospecimen collection.

A four-stage stratified area probability sample design was used in Wave 1 of the PATH Study to select adults and youth ages 12 to 17 from the U.S. civilian, noninstitutionalized population (CNP); an additional "shadow sample" of youth ages 9 to 11 was selected to be interviewed at later waves. At the first stage, a stratified sample of geographical primary sampling units (PSUs) was selected, in which a PSU was a county or group of counties. For the second stage, within each selected PSU, smaller geographical segments were formed and then a sample of these segments was drawn. At the third stage, the sampling frame consisted of the residential addresses located in these segments. The

fourth stage selected adults and youth from the sampled households identified at these addresses, with varying sampling rates for adults by age, race, and tobacco-use status. Adults were sampled in two phases: Phase 1 sampling used information provided by one adult household member in the household screener and Phase 2 sampling used information that the sampled adult provided in the Phase 2 screener at the beginning of the adult interview. Parents do not constitute a separate sample. Parents who provide permission for their child to complete a youth interview are asked to complete a brief parent interview about their youth selected for the PATH Study. All Wave 1 sample participants form the Wave 1 Cohort.

At Wave 4, the Wave 1 Cohort was replenished with a probability sample of adults, youth, and shadow youth ages 10 to 11 selected from the CNP at the time of Wave 4. This sample was selected from residential addresses not selected for Wave 1 in the same sampled PSUs and segments using similar within-household sampling procedures. This "replenishment sample" is designed to supplement the Wave 1 sample. It is not intended to be used for estimation purposes on its own; rather, it was intended to be combined for estimation and analysis purposes with Wave 4 adult and youth respondents from the Wave 1 Cohort who were in the CNP at the time of Wave 4. This combined set of Wave 4 participants forms the Wave 4 Cohort (see Figure 1). Waves in which the PATH Study administered household screeners to sampled adults, youth and shadow youth are herein referred to as recruitment waves.

Computer-assisted personal interviewing (CAPI) was used for the household screener in Wave 1 and Wave 4. CAPI was also used for the parent interviews in each wave. Audio computer-assisted self-interviewing (ACASI) was used for the adult and youth interviews. After Wave 1, there were two annual follow-up waves of the Wave 1 Cohort, referred to as Wave 2 and Wave 3. Wave 4 is the third annual follow-up wave for the Wave 1 Cohort, in addition to the first wave of the Wave 4 Cohort. Table 1 provides the data collection dates and numbers of interviews in each completed wave for which data are available.

Table 1. PATH Study data collection summary, by wave

Wave	Data collection		Interviews conducted*		
	Start date	End date	Adults	Youth	Parents of youth
1	September 12, 2013	December 14, 2014	32,320	13,651	13,588
2	October 23, 2014	October 30, 2015	28,362	12,172	12,129
3	October 19, 2015	October 23, 2016	28,148	11,814	11,807
4	December 1, 2016	January 3, 2018	33,822	14,798	14,709

*For Wave 4 and subsequent waves, the number of interviews available for analysis in the Wave 1 and Wave 4 Cohorts will be smaller than these totals.

In Wave 1, every respondent who completed an adult interview was asked to provide biospecimens. In the waves after Wave 1, participants completing an adult interview for the first time, including newly-sampled adults in Wave 4, were also asked to provide biospecimens. Additionally, participants completing a youth interview in Wave 4 were asked to provide a urine sample. Three more waves of data and biospecimen collection are planned. Waves 5, 6, and 7 will be biennial. Wave 5 will begin in 2018. The multi-wave design allows for the longitudinal assessment of the participants' patterns of use of tobacco products, tobacco exposures, health, and risks for disease. Other aspects being examined in the PATH Study include participants' changes in awareness, knowledge, risk perceptions, and attitudes about current and newly emerging tobacco products.

1.2 Public Use Data Contents

The public use data consist of the following files that will be included for each wave:

- A data file containing adult questionnaire responses, derived variables, corresponding weight variables (for Waves 1 and 2 only), and related external data (i.e., 2010 Census data), referred to as the "adult data file" in the sections below.
- A data file containing youth questionnaire responses, parent questionnaire responses about the youth and the parent, derived variables, corresponding weight variables (for Waves 1 and 2 only), and related external data (i.e., 2010 Census data), referred to as the "youth/parent data file" in the sections below.
- Starting with Wave 3, a set of files containing weight variables, as there are multiple sets of weights available for analysis of the respective adult and youth data.
- Also starting with Wave 3, a reference data file containing derived variables that simplify the definitions of tobacco use variables in follow-up waves, referred to as the "ever/never reference data file" in sections below.

- A master linkage data file is included with each participant's unique identifier (PERSONID) and indicators as to which file(s) in the list above contain data specific to that respondent within a wave and across waves.

The adult data file and the youth/parent data file have similar content for a given wave. They both include the questionnaire items from the interview, and some derived variables such as tobacco-use definitions. The data files are accompanied by annotated instruments for each interview. The annotated instruments provide a link between the variables and the corresponding questions asked in each interview. These documents can serve as references for data users to interpret variables for analyses. (See Chapter 6 for details about the adult data files and youth/parent data files.) In addition to question wording and routing, users are encouraged to read the "Explanatory Notes" section of the appropriate instrument as that section provides information critical to understanding the complex PATH Study instruments.

Chapter 2 of this User Guide contains information about the multi-stage sample design. Chapter 3 describes the data collection methods and Chapter 4 presents the response rates. Chapter 5 discusses weighting procedures, imputation, variance estimation, and selecting the appropriate weights for analysis. Chapter 6 describes the structure of the adult data files and youth/parent data files, including the record identifier, variable naming convention, and variable values and labels. Chapter 7 provides information about linking individual data files included in the release.

2. Sample Design

The PATH Study is a nationally representative, longitudinal cohort study of adults and youth. The PATH Study's design allows for the longitudinal assessment of patterns of use for a spectrum of tobacco products, including initiation, cessation, relapse, and transitions between products, as well as factors associated with use patterns. The study currently follows two cohorts:

- The Wave 1 Cohort - Participants selected using a probability sample at Wave 1 (conducted September 12, 2013 to December 14, 2014) when they were ages 9 years or older. See Section 2.1 for information regarding the selection of the Wave 1 Cohort.
- The Wave 4 Cohort - Participants selected using a probability replenishment sample at Wave 4 (conducted December 1, 2016 to January 3, 2018) when they were ages 10 years or older combined with members of the Wave 1 Cohort who completed an interview at Wave 4 and were in the CNP at that time. See Section 2.3 for information regarding the selection of the additional participants contributing to the Wave 4 Cohort.

Figure 1 illustrates the relationship between the two cohorts. Table 2 shows the sample sizes for the two cohorts.

Figure 1. Illustration of the relationship between the Wave 1 Cohort and the Wave 4 Cohort

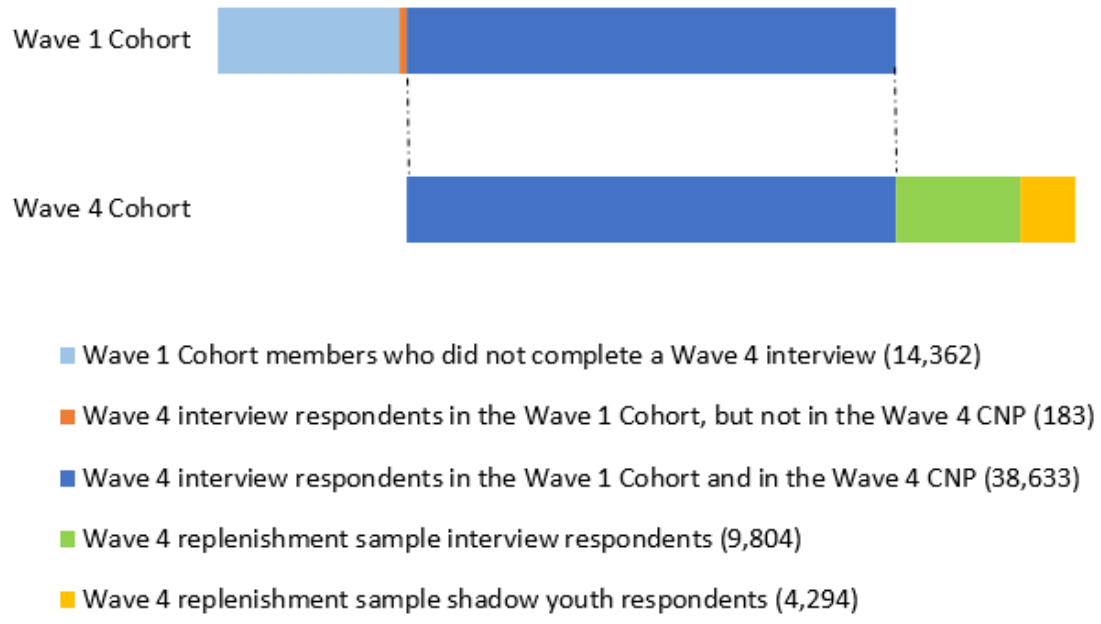


Table 2. Wave 1 Cohort and Wave 4 Cohort Sample Sizes

Sample	Wave 1 Cohort	Wave 4 Cohort
Wave 1 Cohort members who did not complete a Wave 4 interview	14,362	
Wave 4 interview respondents in the Wave 1 Cohort, but not in the Wave 4 CNP	183	
Wave 4 interview respondents in the Wave 1 Cohort and in the Wave 4 CNP	38,633	38,633
Wave 4 replenishment sample interview respondents		9,804
Wave 4 replenishment sample shadow youth respondents		4,294
Total	53,178	52,731

2.1 Wave 1 Sample Design

2.1.1 Target Population

The PATH Study's Wave 1 target population is the civilian household population 9 years of age or older in the United States (all 50 States and the District of Columbia). College students were sampled through their permanent residence rather than at their dormitory. Active-duty members of the military (Army, Navy, Marines, Air Force, and Coast Guard) were excluded, as were all persons living in institutional and noninstitutional group quarters other than college dormitories. Spouses and children of active-duty military living in the 50 States and D.C. were included.

2.1.2 Multi-Stage Sampling

The Wave 1 sample was selected using a four-stage, stratified probability sample design involving the selection of (1) 156 PSUs consisting of counties or groups of contiguous counties, (2) 6,049 second-stage sampling units (referred to as segments), (3) 166,088 mailing addresses, and (4) 76,539 sampled persons within households occupying dwelling units at sampled addresses. The sampling rates for adults varied by age, race, and tobacco-use status. Two-phase sampling was used for adult selection within households to correct for potential misreporting by the household screener respondent of the tobacco-use status of other adult members of the household. The sampling rates for the two phases were designed to achieve sufficiently large sample sizes for young adults (ages 18 to 24) and adult tobacco users of all ages. Generally, up to two youths ages 12 to 17 were sampled within each household. In addition, a "shadow sample" of youth ages 9 to 11 was selected to be interviewed as youth in later waves.

2.1.2.1 PSU Sampling

At the first stage, a stratified sample of 156 PSUs was selected using probability proportional to size (PPS) sampling. The measure of size (MOS) was defined as a weighted sum of estimated PSU adult population counts of the subgroups that would be sampled within households at different rates, where the weights used to construct the MOS were proportional to the expected overall sampling rates to be applied for each subgroup. The use of this composite MOS was designed to (1) give relatively higher probabilities of selection to PSUs with higher proportions of the key subgroups of young adults, Black or African American adults,[1] and tobacco users; (2) to improve the chances that sufficient numbers of the various sampling subgroups would be included in the sample; and (3) to produce more balanced workloads for field interviewers.

Thirty-five PSUs were selected with certainty. These PSUs in effect serve as their own strata for variance estimation. The remaining PSUs were grouped into strata based on known PSU characteristics that were related to the PATH Study variables of interest. This stratification improves the precision of estimates of variables of interest (e.g., tobacco use, perceptions, health, and possible changes over time in those characteristics) because PSUs within the same strata are in general more homogeneous than PSUs in the sampling frame. The variables used for stratification included census region and division, urban/rural designation, Core Based Statistical Area status and size, percent of adults age 25 and older with at least a bachelor's degree, percent of population with family income below 200 percent of the poverty level, percent of population who were Black or African American, and percent of population who were Hispanic. These variables were available at the county level from the 2010 decennial census and the 2006-2010 5-year American Community Survey (ACS) data. Fifty-seven strata were formed using these variables. From each of these 57 strata, two, or occasionally three, PSUs were sampled systematically with probability proportional to the MOS, resulting in a total of 121 non-self-representing PSUs.

2.1.2.2 Segment Sampling

Sampled PSUs were divided into segments based on census-defined blocks and had, in general, a minimum of 100 occupied housing units. This formation of segments sought to achieve desired

[1] Black or African Americans were oversampled because of their higher rate of use of menthol cigarettes compared to other racial/ethnic groups. For this purpose, Black or African American was defined as Black or African American alone or in combination with other race(s), whether Hispanic or non-Hispanic.

sample sizes while minimizing within-segment travel time for field staff. A sample of segments was subsequently selected within each of the sampled PSUs.

Each segment was assigned a MOS to be used in drawing a PPS sample of segments. The segment composite MOS was computed in the same manner as the PSU composite MOS, with the expected counts of adults in each of eight key sampling subgroups defined by age, race, and tobacco-use being computed at the segment level, rather than at the PSU level.

Socio-demographic data from the 2010 Census were used to group and sort the segments on the frame prior to drawing a PPS systematic sample from each PSU. The demographic factors used were percent Black or African American,[2] percent Hispanic, and percent of occupied housing units that were owner-occupied. Segments similar to each other on these three factors were grouped together using a clustering procedure. Consequently, this procedure helped to ensure that a well-balanced sample of segments was taken from each PSU such that the resulting sample of segments would produce estimates with smaller variance (Golder and Yeomans, 1973; Judkins and Singh, 1981). A systematic PPS sample of about 40 segments was drawn within each noncertainty PSU,[3] with more segments drawn in the larger certainty PSUs, for a total of 6,049 segments.

2.1.2.3 Address Sampling

At the third stage of sampling, addresses within each segment were ordered by census block, and a sample was selected by systematic sampling. The goal in assigning address selection probabilities was to minimize the variation in weights across the first three stages of sampling and to produce a relatively even workload for field interviewers in each segment. The sampling rate within each segment was determined so that each address selected for the sample would have approximately the same unconditional probability of selection over all three stages of sampling.[4] Due to the form of the composite MOS used for PPS sampling at the first two stages of the design, the allocated

[2] Black or African American was defined as Black or African American alone or in combination with other race(s), whether Hispanic or non-Hispanic.

[3] A few small PSUs had fewer than 40 segments; in these, all segments were selected with certainty.

[4] The sampling rate was adjusted for some segments to yield an expected number of addresses between 10 and 40. In addition, in some cases, a single address represented a multi-unit structure that contained many dwelling units, all receiving their mail at a single mail "drop point." These drop points were sampled at a higher rate and then a subsample of the dwelling units was drawn from the selected drop point.

number of sampled addresses varied by segment. However, the sampling scheme was intended to produce approximately equal segment workloads in terms of the number of adult interviews conducted. A systematic sample of addresses with the desired sampling rate was chosen from each segment.

2.1.2.4 Within-Household Sampling

The fourth stage of sampling selected persons from the sampled households. During the household screener, one adult household member (referred to as the screener respondent) was asked to list members of the household and provide demographic, and for adults, tobacco-use information about each person for use in sampling three main groups of interest:

- Adults (up to two adults per household were sampled);
- Children ages 12 to 17 (referred to as "youth," generally up to two per household were sampled); and
- Children ages 9 to 11 (referred to as "shadow youth," generally up to two per household were sampled) to be interviewed as youth upon reaching 12 years of age.

The sections below provide more detail regarding the selection of adults, youth, and shadow youth. All adults and youth completing an interview at Wave 1 (45,971 participants), along with all shadow youth permitted by a parent or guardian to participate in the study (7,207 participants), form the Wave 1 Cohort (53,178 participants).

Adults

The sampling procedure for selecting adults within a household had two phases. Phase 1 sampling depended on the information provided by the household screener respondent. For the Phase 1 sampling, adults in the household were classified into one of eight subgroups defined by the cross-classification of age (ages 18 to 24, ages 25 and older), race (Black or African American,[5] all others), and tobacco usage[6] (user, not user), as reported by the screener respondent. Predetermined rates

[5] Black or African American was defined as Black or African American alone or in combination with other race(s), whether Hispanic or non-Hispanic.

[6] Because of the PATH Study's interest in persons who are experimenting with tobacco products or are likely to become users in the future, a broad definition of tobacco use was implemented when classifying adults into Phase 1 sampling

were used to sample adults in Phase 2 of the screening process. This led to the final sampling for the adult interview, subject to the constraint that, at most, two adults could be sampled from each household. Phase 2 sampling allowed for the correction of classification errors by household screener respondents, in particular when the screener respondent erroneously reported that a sampled adult who used tobacco was a nonuser. Phase 2 sampling was based on the sampled individual's self-reported information and thus considered more accurate.[7] The sampling rates for the two phases were designed to achieve predetermined minimum sample sizes for key subgroups, such as young adults (ages 18 to 24) and adult tobacco users of all ages.

At Phase 1, sampling rates for nonusers were kept within reasonable bounds, compared to the sampling rates for users to ensure that the weights of any adults sampled at Phase 1 as nonusers who then reported themselves at Phase 2 to be users would be similar to the weights of those who were correctly classified as users at Phase 1. Misclassification in the other direction—with the screener respondent reporting the adult as a user when the person self-reported as a nonuser—was handled by deselecting some members of this group so that those retained would have sampling rates similar to those of other nonusers.

The population proportions of adults within the eight age/race/tobacco-use subgroups were unknown at the initial design stage. The within-household adult sampling rates were adjusted during Wave 1 data collection as more accurate information was accrued in order to achieve the desired sample sizes in the eight domains.

The two-phase sampling procedure and disproportionate sampling of younger adults, Black or African American adults, and tobacco users had two main effects on the sampling errors. First, the procedure resulted in larger sample sizes in the oversampled subgroups than would have occurred using equal probability sampling, which reduced the sampling errors for estimates calculated for those subgroups. However, the increased precision for those subgroups increased weight variation for the adult sample. The average selection probabilities for subgroups ranged from approximately 0.10 (for non-Black or African American nonusers ages 25 and older) to 0.80 (for Black or African American users ages 18 to 24), which leads to weight factors between 10 and 1.25. Consequently,

domains. This definition classified an adult as a tobacco user if it was reported that he or she smoked a cigarette, cigar, or pipe, or used smokeless tobacco every day or some days; and/or had ever used an e-cigarette, snus, or dissolvable tobacco, or had ever smoked tobacco in a hookah.

[7] Because of the PATH Study's interest in persons who are experimenting with tobacco products or are likely to become users in the future, a "wide net" definition of tobacco use was implemented when classifying adults into Phase 2 sampling domains. This definition classified an adult as a tobacco user if he or she had smoked a cigarette, cigar, or pipe, or had used smokeless tobacco in the past 30 days; and/or had ever used an e-cigarette, snus, or dissolvable tobacco, or had ever smoked tobacco in a hookah.

under this design, estimates calculated for all adults have larger standard errors than they would if all adults were sampled at the same rate. However, sampling all adults at the same rate would have required an enormous sample size to achieve the number of tobacco users needed for the scientific aims of the PATH Study and results in a large group of nonusers, making it difficult to accurately characterize the nature and extent of tobacco product use in the U.S. population.

Youth and Shadow Youth

The PATH Study youth sample has two components: selection of youth ages 12 to 17 for the Wave 1 and follow-up interviews, and selection of a "shadow sample" of youth ages 9 to 11 for inclusion in future waves. Youth in the shadow sample were selected at Wave 1 for the purpose of replenishing the 12- to 17-year-old youth sample in later waves; they were not interviewed at Wave 1. The sampling of youth within sampled households was independent of adult sampling and did not involve any oversampling by race, ethnicity, sex, or tobacco use. The number of persons in the age range 12-17 was tallied for the household. If only one or two children in that age range were in the household, those children were included in the youth sample. If there were more than two children in that age range, two were randomly selected for the youth sample.

The shadow youth selection procedure was identical to that for the youth ages 12 to 17 and was carried out independently of the youth and adult sampling within sampled households. The number of 9- to 11-year-old children in the household was tallied. If only one or two children in that age range were in the household, those children were included in the shadow sample. If there were more than two children in that age range, two were randomly selected for the shadow sample.

Given a special analytic interest in multiple-birth youth, the shadow and youth sampling procedures were modified when households containing multiple-birth youth were encountered so that the multiple-birth youth would have relatively higher probabilities of selection. This resulted in some households with more than two children selected for the youth and/or shadow youth samples.

2.2 Sample Design for Waves 2 and 3

There was no additional sampling for Waves 2 or 3. All Wave 1 participants were eligible for Wave 2 as long as they continued to live in the U.S. and were not incarcerated. For longitudinal studies, issues of eligibility and target populations of inference over time require careful consideration. As

stated in Section 2.1, the PATH Study target population at Wave 1 excluded all active-duty members of the military and all persons living in group quarters other than college dormitories. The exclusion applied to both institutional and noninstitutional group quarters. Thus, the target population for the PATH Study in Wave 2 is the resident population of the U.S. at the time of Wave 2 (other than those who are incarcerated) who were in the CNP 9 years of age or older at Wave 1. Similarly, the target population for the PATH Study in Wave 3 is the resident population of the U.S. at the time of Wave 3 (other than those who are incarcerated) who were in the CNP 9 years of age or older at Wave 1.

2.3 Wave 4 Sample Design

The objectives of the Wave 4 sample design were to replenish the Wave 1 sample to account for the aging of Wave 1 participants and sample attrition. Wave 4 was also an opportunity to replenish the shadow sample as a source of 12- and 13-year-olds in future waves. The replenishment resulted in the formation of a new cohort.

Section 2.3.1 below describes the target populations for the Wave 1 Cohort and the Wave 4 Cohort at the time of Wave 4, respectively. Section 2.3.2 describes the sample selection of the replenishment portion of the Wave 4 Cohort.

2.3.1 Target Populations

For the Wave 1 Cohort, Wave 4 of the PATH Study was the third follow-up wave. These participants were eligible for Wave 4 if they were still residents of the U.S. and not incarcerated. The target population for the Wave 1 Cohort in Wave 4 is the resident population of the U.S. at the time of Wave 4 (other than those who are incarcerated) who were in the CNP 9 years of age or older at Wave 1.

At Wave 4, a probability replenishment sample of adults, youth, and shadow youth (ages 10 and 11) was selected from the U.S. CNP at the time of Wave 4, including persons who were not in the CNP at the time of Wave 1 (such as recent immigrants or those returning from the military). Only members of the Wave 1 Cohort who were also in the CNP at the time of Wave 4 were combined with the Wave 4 replenishment sample to form the new Wave 4 Cohort. As a result, the target

population for the Wave 4 Cohort is the U.S. CNP who are 10 years of age or older in the U.S. (all 50 States and the District of Columbia) at the time of Wave 4.

2.3.2 Replenishment Sampling

As part of the Wave 4 replenishment effort, adults, youth, and shadow youth were sampled within the existing PATH Study sample segments from among the addresses not selected for Wave 1. The goals for the Wave 4 address selection were to achieve desired sample sizes, to minimize variation in weights across the first three stages of sampling, and to produce a relatively even workload for field interviewers in each segment. A total of 174,273 mailing addresses were selected. To meet the needs for the Wave 4 Cohort shadow sample (i.e., to satisfy Wave 5 youth sample size requirements for 12- and 13-year-olds), a randomly selected subset of the sampled addresses (115,536 or close to two-thirds of the addresses) were screened solely to identify shadow youth ages 10 to 11. The remaining addresses (58,737) were screened for adults, youth, and shadow youth ages 10 to 11. These are referred to as the "SO" (shadow youth only) and "AYS" (adults, youth, and shadow youth) replenishment samples, respectively.

The Wave 4 within-household sampling procedures mirrored those used at Wave 1, with sampling rates varying for adults by age, race, and tobacco-use status;[8] however, the within-household sampling rates were designed to bring the Wave 4 adult and youth sample sizes up to a level commensurate with Wave 1 sample sizes by sampling domain. This required consideration for the expected combined effect of the aging of Wave 1 participants and the loss of sample size due to attrition on the PATH Study sample by the time of Wave 4. In particular, it was necessary to oversample adult tobacco users because the Wave 1 youth respondents who became adults since Wave 1 (and those who will continue to do so) were not oversampled with respect to tobacco-use status.

A total of 18,360 adults and youth 12 to 17 were selected as part of the replenishment sample. Those who completed an interview were combined with Wave 1 Cohort interview respondents at Wave 4 in the CNP at that time (48,437 participants total). These interview respondents together with the

[8] The same two-phase sampling procedure for adults used at Wave 1 (and described in Section 2.1.2.4) was used for the replenishment sample, but the definitions of tobacco use were expanded to include users of all e-products and not only users of e-cigarettes.

newly sampled shadow youth permitted by a parent or guardian to participate in the study (4,294 participants) form the Wave 4 Cohort (52,731 participants).

3. Data Collection

The PATH Study collects interview data through in-person CAPI and ACASI methods. Through computer-assisted interviews, the PATH Study collects baseline and follow-up information on tobacco-use patterns, trends in risk perceptions and attitudes regarding harmful constituents, new and emerging tobacco products, and tobacco initiation, cessation, and relapse behaviors among youth ages 12 to 17 and adults ages 18 and older. Parents who provide permission for their child to complete a youth interview are asked to complete a brief parent interview that contains questions about the selected youth's parental supervision, school performance, and tobacco use.

The PATH Study also collects biospecimens from consenting respondents to assess markers of nicotine exposure, to detect and compare intermediate endpoints and incident health outcomes associated with the use of tobacco products and related disease processes, and to validate self-reported behavioral and health data. At each wave, respondents completing the adult interview for the first time are asked to provide samples of urine and blood (and buccal cells, until this collection was discontinued in May 2014). Additionally, some adults who provided urine in a previous wave are asked to continue providing a urine specimen in subsequent waves, with the goal of collecting a specimen from the same individuals over several waves. All youth completing the interview at Wave 4 are also asked to provide a urine specimen.

Data collection at Wave 1 involved four main components: (1) a CAPI household screening instrument, (2) ACASI instruments (separate instruments for youth and adults), (3) a CAPI parent instrument, and (4) collection of urine, blood, and buccal cell specimens from consenting adults.[9] Collection of biospecimens is not a requirement for adult participation; however, completion of an extended interview is required. For Wave 2 and Wave 3, the first component is not necessary. At Wave 4, the CAPI household screener was reinstated and all associated procedures in data and biospecimen collection from Wave 1 were updated to reflect Wave 4 requirements such as the age range of 10 and 11 only for shadow youth. The other data collection components were similar to previous waves as described above.

[9] Buccal cells were collected for a limited time, from September 12, 2013, to May 18, 2014.

Before the start of data collection, the PATH Study received approval from the Westat Institutional Review Board. The PATH Study also obtained a Certificate of Confidentiality from the National Institutes of Health. Amendment 2 of CC-DA-12-131, dated September 23, 2016, extends the certificate's expiration date from September 30, 2017 to August 31, 2022.

The following sections provide more information regarding data collection in each completed wave for which data are available. Table 1 in Chapter 1 details the data collection dates and number of interviews conducted. Information about biospecimen collection is provided to give a full picture of data collection in the PATH Study. Biomarker data are not currently available on the PUFs. However, indicators of biospecimen collection by type of biospecimen are included on the data files.

3.1 Wave 1 Data Collection

3.1.1 Advance Mail and Prepaid Incentives

Advance letters and brochures were mailed to all sampled addresses 1 to 2 weeks prior to the field interviewer's Wave 1 visit to inform the target respondents of the PATH Study sponsor, the nature and uses of the data collected, legal authorities, the voluntary nature of participation, and protection of the information. The advance letter contained a $2 bill as an attention-getter. Before administration of the household screener, the field interviewer asked if the household received the introductory letter and brochure. Those respondents who did not recall receiving or reading them were provided with copies, and were given time to read the information and ask any questions.

3.1.2 Household Screener

Up to two adults and two youths per eligible household were randomly selected using a CAPI screening instrument. One or more additional youth could also be selected among households with multiple-birth youth (e.g., twins). The screener respondent was an adult household member age 18 or older; he/she provided oral consent for the screener. The screener used a full household enumeration process to collect information on age for each reported household member, as well as information on race, active military service status, and tobacco usage for each adult household member. The relationship of all household members to the screener respondent was also collected. In addition to household enumeration information, the household screener respondent's and each

sampled person's telephone numbers were collected to allow re-contact of the household for quality control purposes or to set appointments for interviews if any of the sampled persons or parents were unavailable at the time of the screening. The household's mailing address was also collected for purposes of re-contacting the sampled person(s).

The sampling algorithm for selecting up to two adults and two youths (except in the case of multiple-birth youth) per household was programmed within the CAPI screener software and tested extensively before data collection began.

3.1.3 Interview

The data collection procedures differed for adults and youth. For adults, some questions were asked at the beginning of the interview to serve the purpose of Phase 2 sampling, as discussed in Section 2.1.2.4.

Adult Data Collection

Among other purposes, the household screener (which is also the Phase 1 screener for adults) collected a minimum amount of high-level information about each adult's tobacco usage in order to classify him/her sufficiently for potential selection based on the PATH Study sampling algorithm. The household screener obtained tobacco-use information about all adults from the household screener respondent, and this information could be inaccurate. To obtain more complete and accurate information from an adult sampled through the household screener, a Phase 2 screener was administered to the sampled adult directly to ask a more extensive panel of questions about tobacco usage. The Phase 2 questions were also asked in a private setting using ACASI rather than CAPI.

Following the administration of the Phase 1 screener, if the sampled adult was available and had an adequate amount of time to complete the interview, the field interviewer (1) requested informed consent; (2) administered the Phase 2 screener and adult interview, which included gathering additional contact information about the adult; (3) requested consent for the biospecimen collection; (4) collected the urine specimen (and buccal cell specimen through May 2014); (5) arranged a follow-up appointment for a phlebotomist to collect a blood specimen; and (6) at the completion of the first home visit, paid an incentive to the respondent. If the sampled adult was unavailable or unable

to complete the interview at that time, the field interviewer attempted to schedule an appointment for a return visit or, at a minimum, determined the best time for a return visit.

To begin the interview, the field interviewer provided a brief automated tutorial to the adult on using ACASI and launched the ACASI interview. The first part of the interview process was the Phase 2 screener, which may have confirmed or contradicted the information provided in the Phase 1 (household) screener by the household screener respondent. Depending on the individual's self-reports (e.g., on tobacco usage), the computerized sampling algorithm may have de-selected the respondent and he or she was not asked to complete the remainder of the extended interview. Throughout the interview, the field interviewer provided aid to the sampled person if needed on the use of ACASI or related questions. At the end of the extended interview, the field interviewer collected additional contact information for that person and asked the respondent to consent to providing biospecimens.

Adults who completed the interview or were de-selected based on their responses to the Phase 2 screener received $35 (the adult interview incentive) in appreciation for his/her time for completing the interview, as well as a thank-you letter. Sampled adults who initially declined to participate or were difficult to contact were sent a follow-up refusal conversion letter approximately 3 weeks later.

Youth Data Collection

Following the administration of the household screener, if a parent or guardian of the selected youth was available and had time, the field interviewer (1) requested parent permission for the youth to participate, (2) requested consent for the short parent interview, and (3) administered the CAPI parent interview, which included collecting additional contact information from the parent. If a parent of the sampled youth was unavailable or unable to participate at that time, the field interviewer attempted to schedule an appointment for a return visit or, at a minimum, determined the best time for a return visit. Except for emancipated youth, parental permission was required before conducting the youth interview. The parent who completed a parent interview about the youth received $10 as a token of appreciation for completing the interview.

For a selected youth with parental permission, if the youth was available and had an adequate amount of time to complete the interview, the field interviewer requested youth assent. If a sampled youth was unavailable or unable to complete the interview at that time, the field interviewer

attempted to schedule an appointment for a return visit or, at a minimum, determined the best time for a return visit.

After obtaining assent from the selected youth, the field interviewer provided a brief automated tutorial on using ACASI and launched the ACASI interview. Throughout the interview, the field interviewer provided aid to the sampled youth in using the ACASI instrument if necessary.

Youth who completed the extended interview received $25 (the youth interview incentive) as a token of appreciation for completing the interview. The parent of the youth respondent received a thank-you letter. The parents of sampled youth who were difficult to contact or initially declined permission for their youth to participate were sent a refusal conversion letter approximately 3 weeks later.

3.1.4 Biospecimen Collection

The field interviewer asked each adult interview respondent to consent to provide biospecimens as part of the PATH Study. However, providing biospecimens was voluntary and not a condition of participation. Completion of the Wave 1 adult interview was required from all respondents in order for them to provide biospecimens.

Buccal Cells and Urine Collection

For adults who consented to provide buccal cells and/or urine, the field interviewer collected the specimens following the completion of the interview. The field interviewer provided written and oral instructions to the respondent for collection of buccal cells and/or the urine specimen. The field interviewer packed the specimen(s) and shipped the package to the PATH Study biorepository.

If the adult was unavailable or unable to continue with specimen collection immediately following completion of the interview, the field interviewer attempted to schedule an appointment for a return visit or, at a minimum, determined the best time for a return visit.

Adults who provided a biospecimen during a field interviewer's visit received $25 as a token of appreciation for participating in the buccal cell and/or urine collection component of the PATH Study.

Midway through Wave 1 in May 2014, the PATH Study ended buccal cell specimen collection.

Blood Collection

For adults who consented to provide blood, the field interviewer scheduled the appointment for the visit by a phlebotomist to obtain the blood specimen. Upon visiting the respondent's home, the phlebotomist administered blood suitability exclusion questions for blood collection (CAPI instrument) and asked the respondent to answer items about his/her recent use of tobacco products (ACASI instrument). The phlebotomist then collected the blood specimen, and packed and shipped it to the PATH Study biorepository.

Adults who provided a blood specimen during this later home visit received $25 as a token of appreciation for participating in the blood collection component of the PATH Study.

3.2 Wave 2 Data Collection

Wave 2 was the first annual follow-up wave of the PATH Study. All Wave 1 participants were eligible for Wave 2 as long as they continued to live in the U.S. and were not incarcerated. In order to define a reasonable target period for completing the Wave 2 interview, each participant was assigned an "anniversary month" for Wave 2 (approximately 1 year after the Wave 1 interview was completed) and a target data collection period that would improve the likelihood of the interviews taking place 1 year after each individual's Wave 1 interview and give the field interviewers a target for their data collection efforts.

The Wave 2 anniversary month was defined as the calendar month containing the date of the earliest Wave 1 interview completed by a member of the study participant's household, one year after the Wave 1 interview. Given the challenges of contacting and scheduling the interviews, the target period encompassed 4 months, starting with the month before the anniversary month and ending two months after the anniversary month; however, if necessary, efforts to complete the interview could continue past this period. Data collection efforts for nonrespondents could have continued up to the last day of the data collection period.

3.2.1 Advance Mail and Telephone Contact

As noted in Section 3.1.3, after the Wave 1 interview, adult respondents and parents of responding youth were mailed thank-you letters. They were also mailed cards on the occasion of the adult or youth participant's birthday. These mailings were conducted in order to maintain current contact information and to keep participants and parents engaged in the study.

In addition to these contacts, approximately 3 months in advance of the household anniversary month, a reminder letter was mailed to adult participants, parents of youth participants, and parents of shadow youth to remind them of the PATH Study and the upcoming contact from a field interviewer. Also included in the mailing was a refrigerator magnet, a form for updating contact information, and a postage-paid envelope to encourage participants to confirm or update their contact information. Approximately 1 month prior to the household anniversary month the field interviewer telephoned the adult participant and the parent of the youth participant to arrange a convenient time for an in-person visit by the interviewer to the person's home.

There is one exception to the timing of the above. Parents in households with only shadow youth who were not expected to be 12 years old at Wave 2 were mailed the package 6 months in advance of the anniversary month. This timing helped the PATH Study increase contact and engagement with such parents because they were not as involved in the study as other participants. Again, the package included a refrigerator magnet, a form for updating contact information, and a postage-paid envelope. During the target period, the household was telephoned to obtain updated contact information.

3.2.2 Interview

The data collection procedures differed for: (1) adults at Wave 1 (continuing adults), (2) youth at Wave 1 who were age 17 or younger at Wave 2 (continuing youth) and their parents, (3) youth who turned 18 and were eligible for the adult interview at Wave 2 (aged-up adults), and (4) shadow youth who turned 12 and were eligible for the youth interview at Wave 2 (aged-up youth) and their parents. As in Wave 1, the field interviewer provided a brief automated tutorial on using ACASI to all adult and youth participants, and launched the appropriate automated ACASI interview. The field interviewer was available throughout the interview to assist the participant if needed on the use of ACASI or related questions.

The Wave 2 interviews build on the information collected from adults and youth who completed interviews at Wave 1, i.e., at their baseline. Stable information such as demographic characteristics (e.g., sex and race) was collected only at baseline. Similarly, information on lifetime use of tobacco products up to the time of the baseline interview was not asked again. The Wave 2 interview updated information on the use of the products since the Wave 1 interview.

Data Collection for Continuing Adults

At the in-person visit, the field interviewer: (1) reviewed the main elements of the informed consent for interview provided by the adult upon recruitment; (2) administered the adult interview, which included updating contact information about the adult; (3) as appropriate, reviewed the main elements of consent for the biospecimen collection obtained at recruitment; (4) collected the urine specimen (urine was requested from some adults who provided urine specimens after their first adult interview); and (5) paid the incentive to the respondent at the completion of the first home visit. (The urine collection procedures were the same as described in Section 3.1.4.) If an adult was unavailable or unable to complete the interview at the scheduled time, the field interviewer attempted to schedule an appointment for a return visit or, at a minimum, determined the best time for a return visit.

Adults who completed the interview received $35 (the adult interview incentive) as a thank you for completing the interview. A refusal conversion letter was sent to adults who initially declined to participate or were difficult to contact.

Data Collection for Continuing Youth

At the in-person visit, the field interviewer: (1) reviewed with the parent the main elements of parent permission for the youth to participate obtained upon recruitment; (2) reviewed with the parent the main elements of consent for the short parent interview obtained at recruitment; and (3) administered the CAPI parent interview, which included updating contact information for the parent. If a parent of the youth was unavailable or unable to participate at that time, the field interviewer attempted to schedule an appointment for a return visit or, at a minimum, determined the best time for a return visit. The youth interview was not conducted until parental informed

consent was reviewed. Parents who completed a parent interview for the youth received $10 as a thank you for completing the interview.

If the youth had parental permission, was available, and had an adequate amount of time to complete the interview, the field interviewer reviewed with the youth the main elements of assent for the interview obtained at baseline and requested the youth to complete the automated ACASI instrument. If the youth was unavailable or unable to complete the interview at that time, the field interviewer attempted to schedule an appointment for a return visit or, at a minimum, determined the best time for a return visit.

Youth who completed the interview received $25 (the youth interview incentive) as a thank you for completing the interview. Parents of youth respondents received a thank you letter. A refusal conversion letter also was sent to the parents of youth who were difficult to contact.

Data Collection for Aged-up Adults

At the in-person visit, the field interviewer: (1) requested informed consent; (2) administered the adult interview, which included gathering additional contact information about the adult; (3) requested consent for the biospecimen collection; (4) collected the urine specimen if consent was given; (5) arranged a follow-up appointment for a phlebotomist to collect a blood specimen if consent was given; and (6) paid the incentive to the respondent at the completion of the first home visit. The biospecimen collection procedures for these adults were the same as described in Section 3.1.4. The incentive amounts were the same as described for continuing adults. If an aged-up adult was unavailable or unable to complete the interview at that time, the field interviewer attempted to schedule an appointment for a return visit or, at a minimum, determined the best time for a return visit.

Data Collection for Aged-up Youth

At the in-person visit, the field interviewer: (1) requested parent permission for the youth to participate; (2) requested consent for the short parent interview; and (3) administered the CAPI parent interview, which included updating contact information for the parent. If a parent of an aged-up youth was unavailable or unable to participate at that time, the field interviewer attempted to schedule an appointment for a return visit or, at a minimum, determined the best time for a return

visit. The youth interview was not conducted until parental informed consent was obtained. The parent and youth incentive amounts were the same as described for continuing youth.

If the youth had parental permission, was available, and had an adequate amount of time to complete the interview, the field interviewer requested youth assent and, if provided, requested the youth to complete the automated ACASI instrument. If an aged-up youth was unavailable or unable to complete the interview at that time, the field interviewer attempted to schedule an appointment for a return visit or, at a minimum, determined the best time for a return visit.

3.3 Wave 3 Data Collection

Wave 3 was the second annual follow-up wave for the PATH Study. All Wave 1 Cohort members were eligible for the Wave 3 interview as long as they continued to live in the U.S. and were not incarcerated. To define a reasonable target period for completing the Wave 3 interview, each participant was again assigned an "anniversary month" for Wave 3 and a target data collection period that would improve the likelihood of the interviews taking place at 1-year intervals. The anniversary month was approximately 1 year after the Wave 2 interview was completed or, if there was no Wave 2 interview, 2 years after the completion of the Wave 1 interview.

The target period for contacting and scheduling the interviews encompassed 3 months, starting with the month before the anniversary month and ending 1 month after the anniversary month; however, if necessary, efforts to complete the interview could continue past this period. For some nonrespondents, data collection efforts may have continued up to the last day of the data collection period.

3.3.1 Advance Mail and Telephone Contact

The advance mailing and telephone contact leading up to Wave 3 followed the same protocol as that preceding Wave 2 and described above in Section 3.2.1.

3.3.2 Interview

The data collection procedures differed for: (1) participants who had completed an adult interview in any previous wave (continuing adults), (2) participants who had completed a youth interview in any

previous wave and who were still between the ages of 12 and 17 inclusive (continuing youth) and their parents, (3) youth who turned 18 and were eligible for the adult interview at Wave 3 (aged-up adults), and (4) shadow youth who turned 12 and were eligible for the youth interview at Wave 3 (aged-up youth) and their parents. Unlike Waves 1 and 2 in which the ACASI tutorial was required of everyone, a brief automated tutorial on using ACASI in Wave 3 was required of all aged-up youth as they had no prior ACASI experience but was optional for all other adults and youth. The field interviewer set up the ACASI interview and was available throughout the interview to assist the respondent if needed on the use of ACASI or related questions. The interviewer administered the parent interview using CAPI.

The Wave 3 interview builds on the information collected from the study participant in a prior wave. Questions regarding stable information such as demographic characteristics (e.g., sex and race) and information on lifetime use of tobacco products asked in a previous interview were not asked again.

At the in-person visit, the field interviewer followed the same procedures described in Section 3.2.2 with a few exceptions. A refusal conversion letter was sent to adults (continuing or aged-up) and parents who initially refused to participate in Wave 3 with one exception. In Wave 3, if an adult or parent had already refused participation in Wave 2 and refused again in Wave 3, that refusal was considered final. Refusal conversion was not attempted in these cases. If a youth (continuing or aged-up) refused the interview after the parent provided consent, the field interviewer did not attempt to convert that refusal.

3.4 Wave 4 Data Collection

Wave 4 data collection procedures mirrored both those of Wave 1 for the replenishment sample, and Waves 2 and 3 for the Wave 1 Cohort. Like Wave 1, Wave 4 data collection for the replenishment sample included newly sampled addresses and sampled adults, sampled youth and their parents, and shadow youth and their parents.

Wave 4 was the third annual follow-up wave for the Wave 1 Cohort. All Wave 1 participants were eligible for Wave 4 as long as they continued to live in the U.S. and were not incarcerated. However, Wave 1 Cohort members who refused to participate in both Wave 2 and Wave 3 were not approached for an interview at Wave 4.

To ensure operational efficiency, each Wave 1 Cohort member was again assigned an anniversary month as described in Section 3.3. As in previous waves, data were not collected from shadow youth in Wave 4. Any remaining shadow youth from the Wave 1 Cohort were anticipated to be youth (12 years old) by Wave 4; they were interviewed only if the parent confirmed their age as 12 years old or older.

3.4.1 Advance Contact

For the Wave 1 Cohort, the advance mailing and telephone contact leading up to Wave 4 followed the same protocol as that preceding Waves 2 and 3 and described in Section 3.2.1.

For the AYS portion of the replenishment sample, advance letters and brochures were mailed to each sampled address several weeks prior to the field interviewer's first contact. The advance letter contained a $2 bill to gain the recipients' attention.

Sampled addresses for the SO replenishment sample were mailed a paper screener designed to determine the presence of any 10 or 11 year olds in the household as the first step in data collection. A postage-paid envelope, letter of introduction, and a $2 bill were included with the paper screener. Addresses for which the paper screener was not returned were released for in-person data collection at the conclusion of the mail protocol. Households at sampled addresses that returned the paper screener and reported a 10 or 11 year old household member were also released for in-person data collection within two weeks of receiving the returned paper screener. Advanced letters were mailed to thank the household for returning the paper screener and to notify them of the upcoming field interviewer visit. If a paper screener was returned reporting the absence of 10 or 11 year olds, the household at that address was not contacted further.

3.4.2 Household Screener (Replenishment Sample)

The household screener for Wave 1 (see Section 3.1.2) was reinstated for the replenishment sample. The purpose of the screener was to identify eligible household members to participate in the PATH Study and to select specific participants. The same household screener was used for the AYS and SO replenishment samples.

3.4.3 Interview

The data collection procedures differed for: (1) Wave 1 Cohort adults who had previously completed an adult interview (continuing adults), (2) Wave 1 Cohort youth who had previously completed a youth interview and who were still ages 12 to 17 (continuing youth) and their parents, (3) Wave 1 Cohort youth who turned 18 and were eligible for the adult interview at Wave 4 (aged-up adults), (4) Wave 1 Cohort shadow youth who turned 12 and were eligible for the youth interview at Wave 4 (aged-up youth) and their parents, (5) adults selected from the AYS replenishment sample at Wave 4 (new adults), and (6) youth selected from the AYS replenishment sample at Wave 4 (new youth) and their parents. For new adults and youth, the data collection procedures mirrored those described in Section 3.1.3 for Wave 1. The data collection procedures for the Wave 1 Cohort mirrored those described in Section 3.3.2 for Wave 3.

One difference between the data collection procedures for youth at Wave 4 was that for the first time in the PATH Study, all youth (continuing, new, and aged-up) who completed a Wave 4 youth interview were asked to provide a urine sample in Wave 4. Field interviewers requested permission from parents of youth to collect a urine sample. If permission was given, field interviewers administered the assent for urine collection and collected a urine sample from youth. Youth who provided a urine sample received a $25 incentive.

4. Response Rates

Some addresses sampled for the PATH Study could not be located or accessed, others were found to be ineligible (e.g., vacant lots and group quarters), and some eligible households did not complete the household screener. Further, not all sampled persons within eligible households agreed to participate in the PATH Study, and those who were recruited at Wave 1, i.e., those in the Wave 1 Cohort, may not have responded at some or all of the follow-up waves. Section 4.1 presents the calculations of the Wave 1 response rates for the household screener, the adult interview, and the youth interview. Section 4.2 presents the adult interview and youth interview response rates for Wave 2 through Wave 4 for the Wave 1 Cohort. Section 4.3 presents the household screener, adult interview, and youth interview response rates for the Wave 4 AYS replenishment sample.

As described in Chapter 2, the Wave 4 Cohort is comprised of two groups of study members recruited approximately 3 years apart. Wave 4 was the third follow-up attempt for those sampled at Wave 1, whereas members of the replenishment sample were asked to participate in the PATH Study for the first time. Follow-up wave response rates for the Wave 1 Cohort presented in Section 4.2 are conditioning on Wave 1 response for those sampled at Wave 1, whereas response rates for those from the replenishment sample presented in Section 4.3 are conditioning on completion of the Wave 4 household screener. It is clear that the Wave 4 response rates for the two groups of study members comprising the Wave 4 Cohort are conceptually different. For this reason, no attempt was made to compute "blended" response rates for the Wave 4 Cohort at Wave 4. The response propensities of those Wave 4 Cohort adults and youth who were sampled at Wave 1 are reflected in the Wave 1 Cohort response rates[10] and separate response rates were calculated for adults and youth from the Wave 4 replenishment sample. In addition, household screener response rates were computed for the AYS replenishment sample (from which all Wave 4 replenishment sample adults and youth were selected) at Wave 4.

The response rate calculations were based on the formula provided by the Office of Management and Budget in its "Standards and Guidelines for Statistical Surveys" (2006). This formula calls for

[10] Wave 1 Cohort members who were not in the U.S. CNP at the time of Wave 4 contribute to the Wave 1 Cohort response rates but do not belong to the Wave 4 Cohort. Because this group includes a relatively small number of cases (see Chapter 2), their inclusion or exclusion has no meaningful impact on Wave 4 response rates.

calculating response rates as the ratio of the number of completed cases to the number of eligible sample cases.

4.1 Wave 1 Response Rates

This section summarizes the Wave 1 response rates for the household screener, the adult interview, and the youth interview.

The household screener response rate, denoted as RR_{HH}, was calculated using equation 4.1.1:

$$RR_{HH} = C_{HH}/(C_{HH} + N_{HH} + e_{HH} \times U_{HH}) \quad (4.1.1)$$

where

C_{HH} = number of completed cases;

N_{HH} = number of nonresponding cases known to be eligible;

e_{HH} = estimated proportion of nonresponding cases with unknown eligibility that were eligible; and

U_{HH} = number of nonresponding cases with unknown eligibility.

Wave 1 response rates for adults and youth depended on completion of the Phase 1 household screener.[11] The adult interview response rate (conditioning on the completion of the Phase 1 household screener), denoted as RR_A, was calculated as the product of (1) the Phase 2 screener response rate, and (2) the proportion of adults who completed the adult interview among those who completed the Phase 2 screener and were selected for the adult interview, as shown in equation 4.1.2:

$$RR_A = (C_{P2}/(C_{P2} + N_{P2})) \times (C_A/(C_A + N_A)) \quad (4.1.2)$$

where

C_{P2} = number of completed cases for the Phase 2 screener;

N_{P2} = number of nonrespondents to the Phase 2 screener;

[11] At the adult and youth levels, all cases involved in the response rate calculations were eligible.

C_A = number of completed cases for the adult interview; and

N_A = number of nonrespondents to the adult interview.

The youth interview response rate (conditioning on the completion of the household screener), denoted as RR_Y, was calculated using equation 4.1.3:

$$RR_Y = C_Y/(C_Y + N_Y) \quad (4.1.3)$$

where

C_Y = number of completed cases for the youth interview; and

N_Y = number of nonrespondents to the youth interview.

Unweighted response rates were calculated based on actual case counts, while weighted response rates were based on the sums of the Wave 1 basic design weights. At the household level, the basic design weight is the household inverse probability of selection (IPS) weight described in equation 5.1.1.1 in Chapter 5. At the adult level and youth level, the basic design weight is the product of the household IPS weight and the inverse of the within-household probability of selection. The unweighted response rate measures the success of field operations in obtaining responses from the selected sample. The weighted response rate estimates the proportion of the population represented by the sample group that would have responded if they all had been asked to participate in the study, and it provides a measure of the potential impact of nonresponse on the quality of weighted estimates.

Tables 3 through 5 show the unweighted and weighted response rates for the Wave 1 household screener, adult interview, and youth interview. The weighted response rates were 54.0 percent, 74.0 percent, and 78.4 percent for household screener, adult interview, and youth interview, respectively. Response propensities differed across demographic groups defined by age, sex, race/ethnicity, education, and income level. Differential response rates across demographic groups were compensated for by using nonresponse adjustments in the weighting methods, as described in Chapter 5.

Table 3. Wave 1 household screener response rate

Type of interview	C_{HH}: Completed (n)	N_{HH}: Nonresponse known to be eligible (n)	$e_{HH} \times U_{HH}$: Unknown eligible estimated to be eligible (n)	RR_{HH}: Unweighted response rate (%)	RR_{HH}: Weighted response rate (%)
Household screener	79,198	62,332	4,760	54.1	54.0

Table 4. Wave 1 adult interview response rate

	Phase 2 Screener		Adult Interview			
Type of interview	C_{P2}: P2 screener, completed (n)	N_{P2}: P2 screener, nonresponse (n)	C_A: Adult interview, completed (n)	N_A: Adult interview, nonresponse (n)	RR_A: Unweighted response rate (%)	RR_A: Weighted response rate (%)
Adult interview	44,303	14,785	32,320	80	74.8	74.0

Table 5. Wave 1 youth interview response rate

Type of interview	C_Y: Completed (n)	N_Y: Nonresponse (n)	RR_Y: Unweighted response rate (%)	RR_Y: Weighted response rate (%)
Youth interview	13,651	3,800	78.2	78.4

4.2 Response Rates for Wave 2 through Wave 4 for the Wave 1 Cohort

The response rates for Wave 2 through Wave 4 were calculated as the ratio of the number of completed cases to the number of cases eligible for the interview at each wave among Wave 1 participants.

The response rates were calculated separately for the adult interview and youth interview. Respondents were classified as adults or youth according to their age on the date of the interview: persons 18 or older were asked to complete an adult interview and persons ages 12 to 17 were asked to complete a youth interview.

However, a Wave 1 participant who was a nonrespondent in any follow-up wave did not have an interview date for that wave so the latest available date of birth or age information was used to determine his/her age. The age classification date for a nonrespondent in the Wave 1 Cohort was set to the last day of the last month of the target data collection period (see Sections 3.2 through 3.4) or

the final date of data collection for the wave, whichever came earlier. Nonrespondents determined to be in the age range 12 to 17 on the classification date were classified as youth at the wave; those determined to be 18 or older on the classification date were classified as adults.

Both unweighted and weighted rates were calculated. The numbers of cases used in the calculations for the unweighted rates were the actual case counts. The numbers of cases used in the calculations for the weighted rates were the sums of the Wave 1 basic design weights. The basic design weight is the product of the household IPS weight (described in Section 5.1.1) and the inverse of the within-household probability of selection.

Even though the response rates for adults and youth, denoted as *RR*, were calculated separately, they were calculated using the same equations (4.2.1 and 4.2.2) shown below:

$$RR = C/(C + N + e \times U) \quad (4.2.1)$$

$$e = (C + N)/(C + N + I) \quad (4.2.2)$$

where

C = number of completed cases;

N = number of nonrespondents known to be eligible;

e = estimated proportion of nonresponding cases with unknown eligibility who were eligible.

U = number of nonrespondents with unknown eligibility; and

I = number of ineligible cases.

Table 6 provides the unweighted and weighted response rates for the Wave 2 adult and youth interviews among Wave 1 participants. The weighted response rates were 83.2 percent for adults and 87.3 percent for youth.

Table 6. Wave 2 unweighted and weighted interview response rates for the Wave 1 Cohort

Type of interview	C: Completed (n)	I: Ineligible (n)	N: Nonresponse known to be eligible (n)	U: Nonresponse with unknown eligibility (n)	RR: Unweighted response rate (%)	RR: Weighted response rate (%)
Adult interview	28,362	284	4,266	1,647	82.8	83.2
Youth interview	12,172	20	1,331	444	87.3	87.3

Table 7 provides unweighted and weighted response rates for the Wave 3 adult and youth interviews among Wave 1 participants. Only those participants temporarily, but not permanently, ineligible at Wave 2 (e.g., incarcerated) were included in the calculation of equation 4.2.2. There were a small number of participants discovered to be permanently ineligible at Wave 2 (e.g., deceased, or permanently moved outside the U.S.). This set of cases, denoted as *IP*, is not included in the calculation of equation 4.2.2 but is shown in Table 7 for reference. The weighted Wave 3 response rates for the Wave 1 Cohort were 78.4 percent for adults and 83.3 percent for the youth.

Table 7. Wave 3 unweighted and weighted interview response rates for the Wave 1 Cohort

Type of interview	C: Completed (n)	IP: Permanently ineligible prior to Wave 3* (n)	I: Other ineligible (n)	N: Nonresponse known to be eligible (n)	U: Nonresponse with unknown eligibility (n)	RR: Unweighted response rate (%)	RR: Weighted response rate (%)
Adult interview	28,148	187	436	5,192	2,781	78.0	78.4
Youth interview	11,814	4	27	1,742	636	83.3	83.3

*These cases were not included in equation 4.2.2.

Table 8 provides unweighted and weighted response rates for the Wave 4 adult and youth interviews among Wave 1 participants. Only those participants who had not become permanently ineligible prior to Wave 4 were included in the calculation of equation 4.2.2. The numbers of participants discovered to be permanently ineligible prior to Wave 4 are shown in Table 8 for reference. The weighted Wave 4 response rates for the Wave 1 Cohort were 73.5 percent for adults and 79.5 percent for the youth.

Table 8. Wave 4 unweighted and weighted interview response rates for the Wave 1 Cohort

Type of interview	C: Completed (n)	IP: Permanently ineligible prior to Wave 4* (n)	I: Other ineligible (n)	N: Nonresponse known to be eligible (n)	U: Nonresponse with unknown eligibility (n)	RR: Unweighted response rate (%)	RR: Weighted response rate (%)
Adult interview	27,757	400	519	4,133	6,292	72.9	73.5
Youth interview	11,059	8	36	1,365	1,494	79.5	79.5

*These cases were not included in equation 4.2.2.

Response rates for Wave 2 through Wave 4 differed across groups defined by age, sex, race/ethnicity, and tobacco-use based on self-reported data from the Wave 1 interviews. The weighting adjustments attempted to compensate for differential rates across groups, as described in Chapter 5.

4.3 Response Rates for the Wave 4 Adult, Youth and Shadow Youth (AYS) Replenishment Sample

This section summarizes the household screener, adult interview, and youth interview response rates for the AYS replenishment sample in Wave 4. The household screener response rate, adult interview response rate, and youth interview response rate were calculated using equations (4.1.1), (4.1.2), and (4.1.3), respectively.

Unweighted response rates were calculated based on actual case counts, while weighted response rates were based on the sums of the Wave 4 basic design weights described in Section 5.1.4.3.1 in Chapter 5. At the adult level and youth level, the basic design weight is the product of the household IPS weight and the inverse of the within-household probability of selection. Although it was not designed to generate estimates on its own, the Wave 4 replenishment sample was a probability-based nationally representative sample, so the weighted response may be interpreted as described in Section 4.1.

Tables 9 through 11 show the unweighted and weighted response rates for the Wave 4 AYS replenishment sample. The weighted response rates were 52.8 percent, 68.0 percent, and 70.6 percent for household screener, adult interview, and youth interview, respectively. Response propensities differed across demographic groups defined by age, sex, race/ethnicity, education, and

income level. Differential response rates across demographic groups were compensated for by using nonresponse adjustments in the weighting methods, as described in Chapter 5.

Table 9. Household screener response rate for Wave 4 AYS replenishment sample

Type of interview	C_{HH}: Completed (n)	N_{HH}: Nonresponse known to be eligible (n)	$e_{HH} \times U_{HH}$: Unknown eligible estimated to be eligible (n)	RR_{HH}: Unweighted response rate (%)	RR_{HH}: Weighted response rate (%)
Household screener	27,364	22,010	2,230	53.0	52.8

Table 10. Adult interview response rate for Wave 4 AYS replenishment sample

	Phase 2 Screener		Adult Interview			
Type of interview	C_{P2}: P2 screener, completed (n)	N_{P2}: P2 screener, nonresponse (n)	C_A: Adult interview, completed (n)	N_A: Adult interview, nonresponse (n)	RR_A: Unweighted response rate (%)	RR_A: Weighted response rate (%)
Adult interview	8,927	4,120	6,065	12	68.3	68.0

Table 11. Youth interview response rate for Wave 4 AYS replenishment sample

Type of interview	C_Y: Completed (n)	N_Y: Nonresponse (n)	RR_Y: Unweighted response rate (%)	RR_Y: Weighted response rate (%)
Youth interview	3,739	1,574	70.4	70.6

5. Weights and Imputation

Analysis of data from complex sample designs, such as the PATH Study design, requires the use of weights to compensate for variable probabilities of selection, differential nonresponse rates, and possible deficiencies in the sampling frame (e.g., undercoverage of certain population groups). It is also necessary to implement variance estimation procedures that appropriately account for sampling design factors (such as the stratification and sampling of PSUs and area segments, and the use of oversampling) and nonresponse adjustment factors. Section 5.1 describes the procedures used to calculate the adult and youth weights. Section 5.2 describes methods that should be used to estimate variances. The weights and variance estimation methods permit correct inferences for analyses of the PATH Study adult data and youth/parent data. Section 5.3 provides a description of the imputed variables included in the adult data and youth/parent data. Section 5.4 provides guidelines for selecting the appropriate weights for analysis.

5.1 Adult and Youth Weights

The sections below describe the procedures used to calculate the adult and youth weights for each wave and cohort. The weights variables for each wave, cohort, weight type, and interview type are summarized in Table 12.

Table 12. Weight variables for each wave, cohort, weight type, and interview type*

Wave	Cohort	Weight type	Interview type	Variable name
1	Wave 1	Cross-sectional	Adult Youth	R01_A_PWGT R01_Y_PWGT
2	Wave 1	Single-wave	Adult Youth	R02_A_PWGT R02_Y_PWGT
3	Wave 1	All-waves	Adult Youth	R03_A_AWGT R03_Y_AWGT
3	Wave 1	Single-wave	Adult Youth	R03_A_SWGT R03_Y_SWGT
4	Wave 1	All-waves	Adult Youth	R04_A_A01WGT R04_Y_A01WGT
4	Wave 1	Single-wave	Adult Youth	R04_A_S01WGT R04_Y_S01WGT
4	Wave 4	Cross-sectional	Adult Youth	R04_A_C04WGT R04_Y_C04WGT

*Weight variables are included on the files with the questionnaire data for Wave 1 and Wave 2. For all other waves, these variables are on separate files corresponding to the respective wave, cohort, weight type, and interview type.

Researchers are encouraged to review Section 5.4 for guidelines on selecting the appropriate weight for their analysis.

5.1.1 Wave 1 Weights

The sections below describe the computation of the household weights and the adult and youth weights included on the Wave 1 adult data file and youth/parent data file, respectively.

Household Weights

The initial household weights, denoted as $W1HIPSWT_{ijk}$, were calculated for all sampled addresses as the inverse of the probability of selection (IPS) as shown in equation 5.1.1.1:

$$W1HIPSWT_{ijk} = \frac{1}{P_{ijk}} \quad (5.1.1.1)$$

where P_{ijk} is the probability that household k in segment j of PSU i was selected to be in the sample. However, some sampled addresses could not be located/accessed, others were found to be ineligible (e.g., vacant lots and group quarters), and some eligible households did not complete the household screener. Adjustments were therefore made to the IPS weights of addresses with known eligibility status to compensate for those with unknown eligibility status. This eligibility adjustment was done separately for each census region. The household weight after this adjustment, represented as $W1HNRUNK_{ijk}$, was computed as shown in equation 5.1.1.2:

$$W1HNRUNK_{ijk} = W1HIPSWT_{ijk} \times W1HN_r \quad (5.1.1.2)$$

where $W1HN_r$ is the sum of $W1HIPSWT_{ijk}$ for all selected addresses in census region *r* divided by the sum of $W1HIPSWT_{ijk}$ for addresses with known eligibility status in that region.

Further adjustments were made within weighting classes (or nonresponse adjustment cells) based on information available for both responding and nonresponding households, including census 2010 and ACS 5-year (2009-2013) data pertaining to the segments, tracts, and blocks in which they were located. Census 2010 data were used to calculate the percent of occupied housing units that were

owner-occupied, the percent of the population who were Black or African American,[12] the percent of the population who were Asian,[13] the percent of the population who were Hispanic, and the percent of the population 25 years and older in the census block containing the address. The 5-year ACS data were used to estimate the median monthly housing unit costs in the census tract containing the address. Census region, the urbanicity of the PSU, and the urbanicity of the segment were also used when forming the weighting classes.

A second adjustment was made to account for nonresponse among addresses corresponding to eligible households. Within a weighting class, the weights from the previous step for the responding households were inflated proportionately so that they produced the same sum as the sum of these weights for the responding and nonresponding households combined. The nonresponse-adjusted household weight (denoted as $W1HNRWT_{ijk}$) for responding household k of PSU i and segment j was calculated as shown in equation 5.1.1.3:

$$W1HNRWT_{ijk} = W1HNRUNK_{ijk} \times W1HN_c \quad (5.1.1.3)$$

where $W1HN_c$ is the sum of $W1HNRUNK_{ijk}$ for eligible sampled households in weighting class c of which household k is a member, divided by the sum of $W1HNRUNK_{ijk}$ for all responding households in that weighting class.

The nonresponse-adjusted weights were raked to household counts from the 2013 ACS 1-year Public Use Microdata Sample (ACS PUMS) by census region and household composition. Household composition was defined by the number of non-adult persons in the household (0, 1, or 2+) and the number of adult household members (1, 2, 3+). For raking purposes, the household composition was imputed for households missing this information using logical imputation.[14] The final raked household weight was calculated as shown in equation 5.1.1.4:

$$W1HRKWT_{ijk} = W1HNRWT_{ijk} \times W1HR_{ijk} \quad (5.1.1.4)$$

[12] Black or African American was defined as Black or African American alone or in combination with other race(s), whether Hispanic or non-Hispanic.

[13] Asian was defined as Asian alone.

[14] See Lohr (2010) for a brief description of raking and imputation methods.

where $W1HR_{ijk}$ is the household raking adjustment factor for household k of PSU i and segment j. The household weights are not included on the data files.

Adult Weights

The raked household-level weight was used as the foundation for calculating the adult weight. The adult base weight, denoted as $W1AP1BWT_{ijkl}$, was computed as the product of the final household weight $W1HRKWT_{ijk}$and the reciprocal of the within-household probability of selection for adult l within household k of PSU i and segment j, as shown in equation 5.1.1.5:

$$W1AP1BWT_{ijkl} = W1HRKWT_{ijk} \times \frac{1}{\text{Probability adult } l \text{ selected at Phase 1 from household } (ijk)} \quad (5.1.1.5)$$

The final weights for adults were computed in three steps.

First, a nonresponse adjustment was performed to account for nonresponse to the Phase 2 screener using a combination of census 2010 and 5-year (2009-2013) ACS data (used for the household nonresponse adjustment) and person-level data collected during the household screener. Weighting classes were formed based on census region and urbanicity of the PSU and segment; the percent of occupied housing units that were owner-occupied, the percent of the population who were Black or African American, the percent of the population who were Asian, the percent of the population who were Hispanic, and the percent of the population 25 years and older in the census block containing the address; the median housing unit costs in the census tract containing the address; the age and sex of the household screener respondent; the number of adults in the household (capped at five); and the age, race/ethnicity, sex, and tobacco-use status of the selected adult (as reported by the household screener respondent).[15]

The resulting adult weight (denoted as $W1AP1NRWT_{ijkl}$), adjusted for nonresponse between Phases 1 and 2 of the adult sampling procedure, was calculated for respondents to the Phase 2 Screener as shown in equation 5.1.1.6:

$$W1AP1NRWT_{ijkl} = W1AP1BWT_{ijkl} \times W1AP1N_c \quad (5.1.1.6)$$

[15] Block-level population percentages were created from census 2010 data. Tract-level median housing unit costs were extracted from the 5-year (2009-2013) ACS data.

where $W1AP1N_c$ is the sum of $W1AP1BWT_{ijkl}$ for adults sampled at Phase 1 in weighting class c of which respondent l is a member, divided by the sum of $W1AP1BWT_{ijkl}$ for all adults responding to the Phase 2 screener in that weighting class.

Second, the probability of selection at Phase 2 was used to compute the Phase 2 weight, denoted as $W1AP2WT_{ijkl}$, as shown in equation 5.1.1.7:

$$W1AP2WT_{ijkl} = \mathrm{W1}AP1NRWT_{ijkl} \times \frac{1}{\text{Probability adult } l \text{ from household } (ijk) \text{ selected at Phase 2}}. \quad (5.1.1.7)$$

Finally, raking and trimming were performed in an iterative process. The Phase 2 adult weights were raked to independent population totals based on data from the 2013 ACS PUMS. The raking was done using combinations of census region, age, race/ethnicity, sex, and educational attainment. These variables were imputed if they were missing. (See Section 5.3 for more information about this imputation.)

However, the raking algorithm did not place any restrictions on the highest and lowest values of the raked weights, and a few of the raked weights became extremely large in the process of matching to the population totals from the ACS. To reduce any extreme weights generated in the raking process, a trimming step was performed to bring any extreme weights down to the median weight plus four times the interquartile range[16] within groups defined by the eight age, race, and tobacco-use status categories used to select the adults at Phase 2 (see Section 2.1.2.4). After trimming, the weighted totals no longer matched the control totals, so the raking process was repeated. The trimming and raking steps were iterated until the resulting weights summed to the 2013 ACS PUMS totals for the raking dimensions and were within the bounds defined by the interquartile range criterion.

After the iterative raking and trimming process, the final adult weight, denoted as $W1ARKWT_{ijkl}$, was calculated as shown in equation 5.1.1.8:

$$W1ARKWT_{ijkl} = \mathrm{W1}AP2WT_{ijkl} \times W1ART_{ijkl} \quad (5.1.1.8)$$

where $W1ART_{ijkl}$ is the combined raking and trimming adjustment for adult l within household k of PSU i and segment j. These final weights can be found on the Wave 1 adult data file in the

[16] The weight trimming procedure reduces the mean squared error by reducing the variation among the weights. See Battaglia et al. (2013) and Chowdhury et al. (2007).

variable named R01_A_PWGT, indicating that these are person-level, adult weights for Wave 1 of the PATH Study.

Youth Weights

The raked household-level weight was also used as the foundation for calculating the youth weight. The youth base weight, denoted as $W1YBWT_{ijkl}$, was computed as the product of the final household weight $W1HRKWT_{ijk}$and the reciprocal of the within-household probability of selection for youth l within household k of PSU i and segment j, as shown in equation 5.1.1.9:

$$W1YBWT_{ijkl} = W1HRKWT_{ijk} \times \frac{1}{\text{Probability youth } l \text{ selected from household } (ijk)}. \quad (5.1.1.9)$$

Like the adjustment for household screener nonresponse, a nonresponse adjustment was performed to account for nonresponse to the youth interview using a combination of census 2010 and 5-year (2009-2013) ACS data (used for the household nonresponse adjustment) and person-level data collected during the household screener. Weighting classes were formed based on census region and urbanicity of the PSU and segment; the percent of occupied housing units that were owner-occupied, the percent of the population who were Black or African American, the percent of the population who were Asian, the percent of the population who were Hispanic, and the percent of the population 25 years and older in the census block containing the address; the median housing unit costs in the census tract containing the address; the age and sex of the household screener respondent; the number of adults in the household (capped at five); and the age, sex, and race/ethnicity of the selected youth (as reported by the household screener respondent).

Within a weighting class, the base weights ($W1YBWT_{ijkl}$) for the responding youth were inflated proportionately so that they produced the same sum as the sum of the base weights of the responding and nonresponding youth combined. The nonresponse-adjusted weight for responding youth is shown in equation 5.1.1.10:

$$W1YNRWT_{ijkl} = W1YBWT_{ijkl} \times W1YN_c \quad (5.1.1.10)$$

where $W1YN_c$ is the sum of $W1YBWT_{ijkl}$ for sampled youth in weighting class c of which respondent l is a member, divided by the sum of $W1YBWT_{ijkl}$ for all responding youth in that weighting class.

For youth, the nonresponse-adjusted weights ($W1YNRWT_{ijkl}$) were raked to population totals from the 2013 ACS PUMS using the same iterative process as used in the adult weighting described above. The raking was done using census region, single-year of age, race/ethnicity, and sex as raking variables. These variables were imputed if they were missing. (See Section 5.3 for more information about this imputation.) The trimming threshold was set to the median weight plus four times the interquartile range within groups defined by whether or not the youth was selected with certainty.

After the iterative raking and trimming process, the final youth weight, denoted as $W1YRKWT_{ijkl}$, was calculated as shown in equation 5.1.1.11:

$$W1YRKWT_{ijkl} = \mathrm{W1}YNRWT_{ijkl} \times W1YRT_{ijkl} \qquad (5.1.1.11)$$

where $W1YRT_{ijkl}$ is the combined trimming and raking adjustment for youth l within household k of PSU i and segment j. These final weights can be found on the Wave 1 youth/parent data file in the variable named R01_Y_PWGT, indicating that these are person-level, youth weights for Wave 1 of the PATH Study.

Shadow Youth Weights

In preparation of the shadow youth completing a youth interview in later waves, Wave 1 weights were created for the shadow youth sample so that Wave 1 shadow youth would have a basis for the future-wave weights. These weights are not included with the data files.

The weighting process for the Wave 1 shadow youth was almost identical to that for the Wave 1 youth sample as described above. There were some slight differences, however. These include the variables used to form the weighting classes for the nonresponse adjustment and the level of information available for the raking adjustment.

Nonresponse adjustment was performed to account for nonresponse to the parent consent using a combination of census 2010 and 2009-2013 ACS data (used for the household nonresponse adjustment) and person-level data collected during the household screener. Weighting classes were formed based on census region; the percent of occupied housing units that were owner-occupied, the percent of the population who were Black or African American, the percent of the population who were Hispanic, and the percent of the population 25 years and older in the census block

containing the address; the median housing unit costs in the census tract containing the address; the age and sex of the household screener respondent; the number of adults in the household (capped at five); and the age and race/ethnicity of the selected youth.

Nonresponse-adjusted weights were raked to population totals from the 2013 ACS PUMS using census region, single-year of age, race/ethnicity, and sex as raking variables. However, because no interview is conducted for shadow youth until they reach 12 years of age, the only information available for this adjustment was from the household screener. These variables were imputed if they were missing, as described in Section 5.3, but without the benefit of self-reported information.

5.1.2 Wave 2 Weights

The final weights assigned to the Wave 1 Cohort served as the initial ("base") weights for use in developing the Wave 2 weights. These weights were then adjusted to account for nonresponse to the Wave 2 interview and the resulting weights were raked to Wave 1 population totals (also known as "control totals") as described in the sections below. The weighting process consisted of partitioning the sample into groups defined by Wave 1 age (Section 5.1.2.1), forming weighting classes and performing two nonresponse adjustments (Section 5.1.2.2), and raking to Wave 1 control totals to form the final Wave 2 weights (Section 5.1.2.3).

5.1.2.1 Partitioning the Sample by Wave 1 Age

The nonresponse adjustment was done separately for each of eleven age groups based on Wave 1 age: 9, 10, 11, 12, 13, 14, 15, 16, 17, 18-24, and 25+. The main rationale for doing so was to appropriately handle those who age into a different age group (from shadow youth to youth or from youth to adult). For example, because of the timing of data collection it was possible that some 16-year-olds at Wave 1 (those who were close to 17 at the time of their Wave 1 youth interview) were 18 at Wave 2 because more than 1 year elapsed between interviews, while some 17-year-olds at Wave 1 did not age up to the adult group at Wave 2 because less than a full year had elapsed between interviews. This issue also affected 10- and 11-year-olds regarding turning 12 at Wave 2.

5.1.2.2 Nonresponse Adjustment

Final Wave 1 weights were adjusted for Wave 2 nonresponse in two stages, first accounting for nonresponse among nonrespondents for whom eligibility status could not be ascertained and then among those for whom it could. The first adjustment for nonrespondents whose eligibility status was unknown was undertaken for those in the Wave 1 Cohort who could not be contacted to complete the Wave 2 interview. The second adjustment was undertaken to account for nonrespondents known to be eligible for the Wave 2 interview. This group of nonrespondents largely consisted of those who were contacted but did not complete the Wave 2 interview (e.g., "refusals").

A number of variables were considered in establishing weighting classes for nonresponse adjustment purposes, including variables from census 2010 and 5-year (2009-2013) ACS data (used for the household nonresponse adjustment described in Section 5.1.1), person-level data collected during the Wave 1 household screener, and data from the Wave 1 adult and youth interviews. Because different information was collected in the Wave 1 adult and youth interviews, and no interview data were collected at Wave 1 for the shadow youth, different sets of variables were considered in the formation of the weighting classes for the different sets of Wave 1 participants: one set of variables was considered for adult respondents in Wave 1, another for youth respondents, and a third, smaller set, for shadow youth. The variables comprising the final weighting classes c_U (for the unknown eligibility adjustment) and c_K (for the nonresponse adjustment among those known to be eligible) were determined separately for the eleven age groups based on their ability to discriminate response propensities.

Suppose Wave 1 Cohort member l was assigned weight $W1_l$ for Wave 1 and was assigned to class c_U for the nonresponse adjustment to account for those with unknown eligibility. Then the Wave 2 weight after this adjustment, represented as $W2NRUNK_l$, was computed as shown in equation 5.1.2.1:

$$W2NRUNK_l = W1_l \times W2N_{c_U} \quad (5.1.2.1)$$

where $W2N_{c_U}$ is the sum of $W1_l$ for all Wave 1 Cohort members in weighting class c_U, divided by the sum of $W1_l$ for all Wave 1 Cohort members with known eligibility status at Wave 2 in that weighting class.

This weight was then adjusted for the nonresponse of those known to be eligible for the Wave 2 interview. Let c_K represent the weighting class to which Wave 1 Cohort member l was assigned for this second nonresponse adjustment. Then the final Wave 2 nonresponse-adjusted weight for respondent l, represented as $W2NRWT_l$, was computed as shown in equation 5.1.2.2:

$$W2NRWT_l = W2NRUNK_l \times W2N_{c_K} \quad (5.1.2.2)$$

where $W2N_{c_K}$ is the sum of $W2NRUNK_l$ for all those eligible for Wave 2 in weighting class c_K, divided by the sum of $W2NRUNK_l$ for all Wave 2 respondents in that weighting class.

5.1.2.3 Raking Adjustment

Finally, raking to control totals using Wave 1 characteristics and trimming were performed in an iterative process. Raking to control totals from the recruitment wave is used with longitudinal weights because the target population for longitudinal weights is the target population of initial interest (for the Wave 1 Cohort of the PATH Study, the CNP at the time of Wave 1) followed over time, including those who may enter the military or group quarters (and are not incarcerated) after Wave 1 (see Section 2.2). "Drifting" from important Wave 1 characteristics of analytic interest can arise over time due to nonresponse as well as to Wave 1 Cohort members becoming ineligible in later waves (i.e., some will have died, while others will have moved to a location outside of the U.S.). To maintain consistency with such Wave 1 characteristics, the Wave 2 weights for everyone other than Wave 2 nonrespondents were raked back to control totals for all Wave 1 participants using Wave 1 characteristics. Some of these control totals were simply the population-based control totals used in the Wave 1 raking. Others were sample-based control totals, reflecting PATH Study estimates related to tobacco use.

For Wave 2 adult respondents who were also adults in Wave 1, the population-based control totals used were the same as those used at Wave 1. The sample-based control totals reflected any tobacco use and e-cigarette use reported at Wave 1 cross-classified with some of the same characteristics used to create the population-based control totals (e.g., sex and age) at Wave 1. The tobacco-use variable had three levels: current established user, ever user but not current established user, and never user. Those with missing data for tobacco use were pooled for raking purposes with those associated with the category "ever user but not current established user." The e-cigarette variable

had two levels: never user and other. Those with missing data for e-cigarette use were pooled with those associated with the category "other."

For Wave 2 respondents who completed the youth interview at Wave 1, the population-based control totals used were the same as those used at Wave 1. The sample-based control totals were created from Wave 1 estimates of any tobacco use (for respondents who were 16 or 17 at Wave 1) and e-cigarette use (for respondents who were 17 at Wave 1);[17] the tobacco-use and e-cigarette variables were both defined using two levels: never user and other; those with missing data on either tobacco variable were pooled with those associated with the category "other" for that variable.

For Wave 2 respondents who were shadow youth at Wave 1, the raking was done as described in Section 5.1.1.

However, the raking algorithm did not place any restrictions on the highest and lowest values of the raked weights, and a few of the raked weights became extremely large in the process of matching to the control totals. To reduce any extreme weights generated in the raking process, a trimming step was performed to bring any extreme weights down to the median weight plus four times the interquartile range within groups used to select the PATH Study participants at Wave 1. For those respondents who were adults at Wave 1, these were the eight groups defined by age, race, and tobacco-use status (see Section 2.2.4); for respondents who were not adults at Wave 1, these were groups defined by whether or not they were selected with certainty.

After trimming, the weighted totals no longer matched the control totals, so the raking process was repeated. The final Wave 2 weight, denoted as $W2RKWT_l$, was thus computed for each respondent l as the product shown in equation 5.1.2.3:

$$W2RKWT_l = W2NRWT_l \times W2RT_l \quad (5.1.2.3)$$

where $W2RT_l$ is the combined raking and trimming adjustment for respondent l. The final weight for each Wave 1 Cohort member responding to the Wave 2 adult interview ("A") or youth interview

[17] The ages for which the e-cigarette use and any tobacco use dimensions were used for raking were determined by the need to have sufficient sample size for stability of the weighting adjustments.

("Y") is stored in the variable named R02_*x*_PWGT (where *x* is either "A" or "Y" as appropriate), indicating that it is a person-level weight for Wave 2 of the PATH Study.

5.1.3 Wave 3 Weights

There are two longitudinal weights available for analysis of Wave 3 data: the all-waves weight and the single-wave weight. The Wave 3 all-waves weight was assigned to Wave 3 respondents who participated at both Wave 1 and Wave 2. The Wave 3 single-wave weight was assigned to all Wave 3 respondents whether or not they participated at Wave 2. So participants in all three waves were assigned both the all-waves and single-wave weights. Both weights are longitudinal; however, the single-wave weight could serve as a proxy for a cross-sectional weight.

Section 5.1.3.1 describes the construction of the Wave 3 all-waves weights and Section 5.1.3.2 describes the Wave 3 single-wave weights.

5.1.3.1 Creation of the Wave 3 All-Waves Weights

All-waves weights were created for Wave 3 respondents who also responded at both Wave 1 and Wave 2. This includes Wave 3 respondents who either:

1. Completed an interview (adult or youth) at Wave 1 and Wave 2; or
2. Were in the shadow sample of persons ages 9 to 11 at Wave 1, were at least 12 years old at the time of the Wave 3 interview, and completed an interview for all waves in which they were old enough to do so or verified their information with the study for waves in which they were not old enough to be interviewed.

The Wave 3 all-waves weighting process began with all Wave 2 respondents and the nonresponse-adjusted weights assigned to them ($W2NRWT_l$) during the Wave 2 weighting process. These initial weights were then adjusted to account for nonresponse to the Wave 3 interview as described in the sections below. The nonresponse adjustment process consisted of the following steps: partitioning the sample into groups defined by Wave 2 age and forming nonresponse adjustment cells (Section 5.1.3.1.1), and adjusting for nonresponse and raking to control totals to form the final Wave 3 all-waves weights (Section 5.1.3.1.2).

5.1.3.1.1 Partitioning the Sample by Wave 2 Age and Forming Nonresponse Adjustment Cells

The adjustment of weights for nonresponse was done separately for each of ten age groups based on Wave 2 age, as this was the most recently confirmed age value available for the Wave 2 respondents. These groups were 9- to 10-year-olds and the seven individual ages from 11 to 17 as well as the two adult age ranges 18-24, and 25 and older. This was considered preferable to using the Wave 3 age information which was not confirmed for nonrespondents. Using separate nonresponse adjustment subgroups based on age provides a mechanism for proportionately allocating weights of nonrespondents appropriately between youth and adults or shadow youth and youth.

The initial ("base") weights (the Wave 2 weights reflecting adjustments for nonresponse) were adjusted for Wave 3 nonresponse in two stages. The first stage accounted for nonresponse among the nonrespondents for whom eligibility status could not be ascertained, the second stage among those for whom it could.

There were many variables available for consideration in establishing weighting cells for nonresponse adjustment purposes. One set of variables was considered for those 18 or older (i.e., adults) at Wave 2. Another set was considered for those ages 12-17 (i.e., youth) at Wave 2. A final set was considered for those ages 9-11 (i.e., shadow youth, mostly 10- and 11-year-olds) at Wave 2. The variables comprising the final weighting classes c_U (for the unknown eligibility adjustment) and c_K (for the nonresponse adjustment among those known to be eligible) were determined separately for the ten age groups.

5.1.3.1.2 Computation of the Nonresponse and Raking Adjustments

Suppose a Wave 2 respondent l was assigned weight $W2NRWT_l$ after the Wave 2 nonresponse adjustments and was assigned to cell c_U for the first stage of nonresponse adjustment. Then the resulting Wave 3 all-waves weight after this adjustment, represented as $W3NRUNK\_A_l$, was computed as shown in equation 5.1.3.1:

$$W3NRUNK\_A_l = W2NRWT_l \times W3N\_A_{c_U} \quad (5.1.3.1)$$

where $W3N\_A_{c_U}$ is the sum of $W2NRWT_l$ for all Wave 2 respondents in weighting class c_U, divided by the sum of $W2NRWT_l$ for Wave 2 respondents with known eligibility status at Wave 3 in that weighting class.

This weight was then adjusted for the nonresponse of those known to be eligible for the Wave 3 interview. Let c_K represent the cell to which a Wave 2 respondent l known to be eligible for data collection at Wave 3 was assigned for nonresponse adjustment purposes. Then the Wave 3 all-waves nonresponse-adjusted weight for respondent l, represented as $W3NRWT\_A_l$, was computed as shown in equation 5.1.3. 2:

$$W3NRWT\_A_l = W3NRUNK\_A_l \times W3N\_A_{c_K} \quad (5.1.3.2)$$

where $W3N\_A_{c_K}$ is the sum of $W3NRUNK\_A_l$ for all those eligible for Wave 3 in weighting class c_K, divided by the sum of $W3NRUNK\_A_l$ for all Wave 3 respondents in that weighting class.

Finally, the Wave 3 nonresponse-adjusted weights for everyone other than Wave 3 nonrespondents were raked to control totals using Wave 1 characteristics with trimming also performed as part of an iterative process. The control totals and methods used were the same as those used to create the Wave 2 weights and are described in Section 5.1.2.3; however, some slight modifications were made to the raking dimensions due to small sample sizes.

The final Wave 3 all-waves weight, denoted as $W3RKWT\_A_l$, was computed for each respondent l as shown in equation 5.1.3.3:

$$W3RKWT\_A_l = W3NRWT\_A_l \times W3RT\_A_l \quad (5.1.3.3)$$

where $W3RT\_A_l$ is the combined raking and trimming adjustment for respondent l. The variable representing the final all-waves weight for each Wave 1 Cohort member responding to the Wave 3 adult interview ("A") or youth interview ("Y") is named R03_*x*_AWGT (where *x* is either "A" or "Y" as appropriate). These variables are provided separately from the adult and youth/parent data files on weight files that also contain variables for use in variance estimation (see Section 5.2). Note that only those Wave 3 respondents participating at Wave 1 and Wave 2 will have a Wave 3 all-waves weight (as described at the beginning of Section 5.1.3.1) and corresponding records on these weight files.

5.1.3.2 Creation of the Wave 3 Single-Wave Weights

Wave 3 single-wave weights were assigned to all Wave 3 respondents regardless of their response status at Wave 2. To account for those Wave 1 Cohort members who became permanently ineligible (i.e., died or moved outside the U.S.) between Waves 1 and 2 separately from those who became ineligible between Waves 2 and 3, the Wave 3 single-wave weight was developed in a slightly different fashion from the all-waves weight.

The final Wave 1 weights served as the initial ("base") weights for use in developing the Wave 3 single-wave weights. These initial weights were then adjusted to account for nonresponse to the Wave 3 interview as described in the sections below. The weighting process consisted of partitioning the sample into groups defined by Wave 1 age and forming nonresponse adjustment cells (Section 5.1.3.2.1), and adjusting for nonresponse and raking to control totals to form the final Wave 3 single-wave weights (Section 5.1.3.2.2).

5.1.3.2.1 Partitioning the Sample by Wave 1 Age and Forming Nonresponse Adjustment Cells

When developing the weighting adjustment cells, the same set of data is needed for both nonrespondents and respondents. For the Wave 3 single-wave weighting process, variables available for both Wave 3 respondents and nonrespondents are from Wave 1. As a result, Wave 1 data were used in the formation of weighting adjustment cells after partitioning the sample into eleven groups defined by Wave 1 age (as was done in the Wave 2 weighting process described in Section 5.1.2.1).

As with the Wave 2 and Wave 3 all-waves weighting processes, the nonresponse adjustment was performed in two stages. The variables comprising the final weighting classes c_U (for the unknown eligibility adjustment) and c_K (for the nonresponse adjustment among those known to be eligible) were determined separately for the eleven age groups.

5.1.3.2.2 Nonresponse and Raking Adjustments

Estimating the eligibility status of the nonrespondents whose eligibility at Wave 3 could not be ascertained (because, for example, they could not be located) is a standard part of the nonresponse adjustment process and is handled in a routine fashion for the all-waves weighting, as described in Section 5.1.3.1.2. If the usual "routine" approach were to be used for the single-wave weighting, all

persons known to be ineligible at Wave 3 would be used in determining the weighting adjustments that account for unknown eligibility. Doing so would overstate the portion of nonrespondents of unknown eligibility considered to be ineligible because it would apply an ineligibility rate accumulated across two waves of the study. This would not be a major concern in Wave 3 but could become one in a later wave. Therefore, Wave 1 Cohort members known to be permanently ineligible at Wave 2 were excluded from the Wave 3 single-wave nonresponse adjustment process.

Suppose a Wave 1 Cohort member l received final Wave 1 weight $W1_l$ and was assigned to cell c_U for the Wave 3 single-wave weight adjustment to account for nonrespondents of unknown eligibility. Then the resulting Wave 3 single-wave weight after this adjustment, represented as $W3NRUNK\_S_l$, was computed as shown in equation 5.1.3.4:

$$W3NRUNK\_S_l = W1_l \times W3N\_S_{c_U} \quad (5.1.3.4)$$

where $W3N\_S_{c_U}$ is the sum of $W1_l$ for all Wave 1 Cohort members in weighting class c_U who were not permanently ineligible at Wave 2, divided by the sum of $W1_l$ for Wave 1 Cohort members with known eligibility status at Wave 3 who were not permanently ineligible at Wave 2 in that weighting class.

This weight was then adjusted for the nonresponse of those known to be eligible for the Wave 3 interview. Let c_K represent the cell to which a Wave 1 Cohort member l known to be eligible for data collection at Wave 3 was assigned for nonresponse adjustment purposes. Then the Wave 3 single-wave nonresponse-adjusted weight for respondent l, represented as $W3NRWT\_S_l$, was computed as shown in equation 5.1.3.5:

$$W3NRWT\_S_l = W3NRUNK\_S_l \times W3N\_S_{c_K} \quad (5.1.3.5)$$

where $W3N\_S_{c_K}$ is the sum of $W3NRUNK\_S_l$ for all those eligible for Wave 3 in weighting class c_K, divided by the sum of $W3NRUNK\_S_l$ for all Wave 3 respondents in that weighting class.

After completing the Wave 3 nonresponse adjustment process, the raking process (including ineligibles from Waves 2 and 3 not included in the nonresponse adjustment process) was the same as that employed for the Wave 3 all-waves weighting process described in Section 5.1.3.1.2. The final raked weights of those who responded to a Wave 3 interview were then made available for the analysis of PATH Study data.

The variable representing the final single-wave weight for each Wave 1 Cohort member responding to the Wave 3 adult interview ("A") or youth interview ("Y") is named R03_*x*_SWGT (where *x* is either "A" or "Y" as appropriate). These variables are provided separately from the adult and youth/parent data files on weight files that also contain variables for use in variance estimation (see Section 5.2).

5.1.4 Wave 4 Weights

There are two longitudinal weights available for analysis of Wave 4 data for the Wave 1 Cohort: the all-waves weight and the single-wave weight. The Wave 4 all-waves weight was assigned to Wave 4 respondents who participated at all waves since recruitment. The Wave 4 single-wave weight was assigned to all Wave 4 respondents in the Wave 1 Cohort whether or not they participated at Wave 2 or Wave 3.

In addition, there is a cross-sectional weight for all Wave 4 respondents in the Wave 4 Cohort. The Wave 4 cross-sectional weight was assigned to respondents to the Wave 4 AYS replenishment sample and Wave 4 respondents from the Wave 1 Cohort who were still in the CNP at the time of Wave 4.

Sections 5.1.4.1 and 5.1.4.2, respectively, describe the creation of the Wave 4 all-waves and single wave weights for the Wave 1 Cohort. Section 5.1.4.3 describes the creation of the Wave 4 cross-sectional weights for the Wave 4 Cohort.

5.1.4.1 Creation of the Wave 4 All-Waves Weights for the Wave 1 Cohort

The all-waves weights were created for Wave 4 respondents in the Wave 1 Cohort who also responded at Wave 1, Wave 2, and Wave 3. This includes Wave 4 respondents who either:

1. Completed an interview (adult or youth) at Wave 1, Wave 2, and Wave 3; or
2. Were in the shadow sample of persons ages 9 to 11 at Wave 1, were at least 12 years old at the time of the Wave 4 interview, and completed an interview for all waves in which they were old enough to do so or verified their information with the study for waves in which they were not old enough to be interviewed.

The Wave 4 all-waves weighting process began with all Wave 3 respondents and the nonresponse-adjusted weights assigned to them ($W3NRWT_l$) during the Wave 3 weighting process. These initial weights were then adjusted to account for nonresponse to the Wave 4 interview. Wave 3 interview variables were used in the nonresponse adjustment process, except for participating shadow youth at Wave 3 where the household screener variables were used.

The Wave 3 all-waves weighting process described in Section 5.1.3.1 is applicable to that for Wave 4, substituting "Wave 3" where "Wave 2" has been referenced and "Wave 4" where "Wave 3" has been referenced. The nonresponse adjustment process consisted of the following steps: partitioning the sample into groups defined by Wave 3 age and forming nonresponse adjustment cells, and adjusting for nonresponse and raking to control totals to form the final Wave 4 all-waves weights. The last step of raking to control totals included some slight modifications to the raking dimensions used in previous waves due to small sample sizes.[18]

The variable representing the final all-waves weight for each Wave 1 Cohort member responding to the Wave 4 adult interview ("A") or youth interview ("Y") is named R04_*x*_A01WGT (where *x* is either "A" or "Y" as appropriate). These variables are provided separately from the adult and youth/parent data files on weight files that also contain variables for use in variance estimation (see Section 5.2). Note that only those Wave 4 respondents who also responded at Wave 1, Wave 2, and Wave 3 will have a Wave 4 all-waves weight (as described at the beginning of Section 5.1.4.1) and corresponding records on these weight files.

5.1.4.2 Creation of the Wave 4 Single-Wave Weights for the Wave 1 Cohort

Wave 4 single-wave weights were assigned to all Wave 4 respondents in the Wave 1 Cohort who completed an interview at Wave 1 regardless of their response status at Wave 2 or Wave 3. These weights may be used for longitudinal analyses that use Wave 1 information to estimate an outcome at Wave 4. Because Wave 1 Cohort members selected at Wave 1 as shadow youth do not have any Wave 1 data, and so cannot be used in such an analysis, they were not assigned a Wave 4 single-wave weight. These participants were assigned a single-wave weight at Wave 3 because those weights could be used as to create approximate cross-sectional estimates for that wave. However, genuine

[18] Because of the increased interest in e-cigarette use estimates, the raking dimensions including Wave 1 e-cigarette use were expanded to include 15- and 16-year-olds.

Wave 4 cross-sectional weights were created for the Wave 4 Cohort (as discussed in Section 5.1.4.3), making such approximations unnecessary. See Section 5.4 for more information on selecting the appropriate weights for analysis.

Even though Wave 1 Cohort members selected as shadow youth at Wave 1 were not assigned a final Wave 4 single-wave weight for analysis, they were included in the weighting process so that those participants who are also in the Wave 4 Cohort would have an initial weight for the Wave 4 cross-sectional weighting process (see Section 5.1.4.3). The final Wave 1 weights served as the initial ("base") weights for use in developing the Wave 4 single-wave weights. These initial weights were then adjusted to account for nonresponse to the Wave 4 interview. PATH Study participants known to be permanently ineligible at Wave 3 (including those who became permanently ineligible at Wave 2) were excluded from the Wave 4 nonresponse adjustment process for the single-wave weight.

The Wave 3 single-wave weighting process described in Section 5.1.3.2 is applicable to that for Wave 4, substituting "Wave 3" where "Wave 2" has been referenced and "Wave 4" where "Wave 3" has been referenced. The nonresponse adjustment process consisted of the following steps: partitioning the sample into groups defined by Wave 1 age and forming nonresponse adjustment cells, and adjusting for nonresponse and raking to control totals to form the final Wave 4 single-wave weights. The last step of raking to control totals included some slight modifications to the raking dimensions used in previous waves due to small sample sizes.[19]

The variable representing the final single-wave weight for each Wave 1 Cohort member completing an interview at Wave 1 and either the Wave 4 adult interview ("A") or youth interview ("Y") is named R04_*x*_S01WGT (where *x* is either "A" or "Y" as appropriate). These variables are provided separately from the adult and youth/parent data files on weight files that also contain variables for use in variance estimation (see Section 5.2).

[19] Because of the increased interest in e-cigarette use estimates, the raking dimensions including Wave 1 e-cigarette use were expanded to include 15- and 16-year-olds.

5.1.4.3 Creation of the Wave 4 Cross-Sectional Weights for the Wave 4 Cohort

The target population for the Wave 4 Cohort is the CNP at the time of Wave 4. Thus, the weights for the Wave 4 Cohort are cross-sectional at Wave 4. The approach for forming the cross-sectional weights is as follows:

1. Create weights for the Wave 4 AYS replenishment sample respondents.
2. Partition the set of Wave 4 AYS replenishment sample interview respondents into two groups: those who were members of the Wave 1 CNP and those who were not.
3. For those in the first group identified in step 2, rake their initial weights (from step 1) to population control totals, creating preliminary raked weights.
4. Identify those Wave 1 Cohort members who were both Wave 4 interview respondents and members of the Wave 4 CNP, and assign their Wave 4 single-wave weight after nonresponse adjustment as their initial weight.
5. Rake the Wave 1 Cohort initial weights (from step 4) to the same population control totals used in step 3, creating a second set of Wave 4 respondents with preliminary raked weights.
6. Develop compositing factors for the Wave 1 Cohort and AYS replenishment sample respondents who were members of the Wave 1 CNP (the first group identified in step 2).
7. Apply the compositing factors to the corresponding preliminary raked weights established in steps 3 and 5.
8. Combine (a) all those with composited weights with (b) the set of Wave 4 replenishment sample respondents not included in the compositing process (the second group identified in step 2), along with their initial weight. Together this set of respondents forms the Wave 4 Cohort.
9. Finally, rake and trim the Wave 4 Cohort weights from step 8.

Section 5.1.4.3.1 describes step 1, Section 5.1.4.3.2 describes steps 2-7, and Section 5.1.4.3.3 describes steps 8-9.

5.1.4.3.1 Weighting the Wave 4 AYS Replenishment Sample

The weighting process for the AYS replenishment sample at Wave 4 was similar to that used for the original sample at Wave 1, with the following two exceptions:

First, adjustments made to sampling rates during the Wave 4 field period required the incorporation of two additional weighting factors for adults and one additional weighting factor for youth, compared to Wave 1. These adjustments resulted from the need to monitor and project combined yields for Wave 4 from both the replenishment sample and the Wave 1 Cohort. Further details can be found in the adult and youth weighting subsections below.

Second, the purpose of the replenishment sample was to supplement the Wave 1 Cohort, not to support the development of national estimates based on the replenishment sample alone. As a result, person-level weights for the Wave 4 AYS replenishment sample interview respondents alone were not calibrated to population control totals. Calibration of person-level weights that included all members of the Wave 4 Cohort occurred only as a final step, as described in Section 5.1.4.3.3.

The sections below describe the computation of the household weights as well as the adult and youth weights for the Wave 4 AYS replenishment sample.

Household Weights

The process for creating household weights for the AYS replenishment sample is similar to that described in Section 5.1.1 for Wave 1. Base weights were calculated for all sampled addresses as the inverse of the probability of selection. Adjustments were then made to the base weights to account for sampled addresses whose residential and occupancy status was unknown.

This eligibility adjustment was performed separately within weighting classes. The weighting classes were based on information available for all sampled addresses, including census 2010 and ACS 5-year (2009-2013) data pertaining to the segments, tracts, and blocks in which they were located. Census 2010 data were used to calculate the percent of occupied housing units that were owner-occupied, the percent of the population who were Black or African American, the percent of the population who were Asian, the percent of the population who were Hispanic, and the percent of the population 25 years and older in the census block containing the address. The 5-year ACS data were used to estimate the median monthly housing unit costs in the census tract containing the address. Census region, urbanicity of the PSU, urbanicity of the segment, an address type indicator,[20]

[20] The address type indicator distinguished between the different sources of mailing addresses for sampling.

and an indicator variable for PSUs that were affected by weather conditions[21] were also used when forming the weighting classes.

A second adjustment was made to account for nonresponse among addresses corresponding to eligible households. This nonresponse adjustment was performed separately within weighting classes that distinguished response propensities between subgroups of households, based on information available for both responding and nonresponding households. The same set of variables for the eligibility adjustment described above was used to develop the nonresponse adjustment cells. The nonresponse-adjusted weights were then raked to estimated numbers of households from the 2016 ACS PUMS representing the four census regions.

Adult Weights

The number of adult respondents from the Wave 4 replenishment sample was monitored throughout the wave to ensure that targeted numbers of respondents would be met for various domains: age group (18-24, 25 and older), race (Black or African American, all others), and tobacco usage (user, not user). This monitoring took into account not only the original targeted numbers for the replenishment sample but also yields from the corresponding domains for the Wave 1 Cohort, the domains being replenished. If the yield for a domain from the Wave 1 Cohort was higher than expected, the sampling rate for the corresponding replenishment sample domain was adjusted to compensate.

At one point in the field period, the Wave 4 projected yields for one particular domain (non-Blacks who were at least 25 years old and nonusers of tobacco[22]) were high enough that no more adults were needed from the replenishment sample. Since two-phase sampling was used to select adults from the replenishment sample, we had Phase 1 domains and Phase 2 domains. Phase 1 domains were defined based on the household screener respondent's report of the adult's age, race, and tobacco-use status. Phase 2 domains were defined based on the adult's self-report of these characteristics and whether or not the tobacco-use statuses reported by the household screener

[21] The indicator variable identified clusters of PSUs where data collection efforts were affected by snow/ice storms, hurricanes, mudslides, wildfires, or flooding.

[22] All definitions of tobacco use relating to the domains discussed in this section are based on the "wide net" definition used for sampling adults at Wave 4.

respondent and adult agreed. Partway through data collection, sampling rates were reduced to zero for the following: (1) Phase 1 domain "non-Blacks who were at least 25 years old and nonusers of tobacco," (2) Phase 2 domain "non-Blacks who were at least 25 years old and nonusers of tobacco according to both the household screener respondent and self-report;" and (3) Phase 2 domain "non-Blacks who were at least 25 years old and nonusers of tobacco according to the household screener respondent but users based on self-report." Two adjustment factors were constructed to account for adults in these domains who did not have a chance of selection for the replenishment sample.

The first adjustment factor accounted for those in the Phase 1 domain with no chance of sample selection after the rate changes were implemented. The second adjustment factor accounted for those in the Phase 2 domains with no chance of sample selection after the rate changes were implemented.[23] Application of these adjustment factors served the function of treating the samples that were selected from these domains as subsamples from the full sets of domain members identified over time.

The adjustment factor for adults in domain d_1 based on the Phase 1 screener was computed as shown in equation (5.1.4.1):

$$P1FACTOR_{d_1} = \frac{number\ of\ adults\ (in\ responding\ households)\ in\ d_1}{number\ of\ adults\ in\ d_1\ with\ Phase\ 1\ sampling\ rate > 0}. \qquad (5.1.4.1)$$

For adults in the Phase 1 domain "non-Blacks who were at least 25 years old and nonusers of tobacco" the value of $P1FACTOR_{d_1}$ was 1.469. $P1FACTOR_{d_1}$ was 1 for adults in all other Phase 1 domains.

The adjustment factor for adults in Phase 2 domain d_2 was computed as shown in equation (5.1.4.2):

$$P2FACTOR_{d_2} = \frac{number\ of\ adults\ (responding\ to\ Phase\ 2\ screener)\ in\ d_2}{number\ of\ adults\ in\ d_2\ with\ Phase\ 2\ sampling\ rate > 0}. \qquad (5.1.4.2)$$

[23] No adults had a chance of being in the two Phase 2 domains with zero sampling rates if the household screener was completed after the corresponding Phase 1 domain sampling rate was set to zero. However, there was often a natural delay between the two phases of sampling, so for some adults the rate changes took place after they were selected from the Phase 1 domain and before completing the Phase 2 screener. The Phase 2 adjustment factor addresses this situation.

For adults in the Phase 2 domain "non-Blacks who were at least 25 years old and nonusers of tobacco according to both the household screener respondent and self-report", the value of $P2FACTOR_{d_2}$ was 1.147. For adults in the Phase 2 domain "non-Blacks who were at least 25 years old and nonusers of tobacco according to the household screener respondent but users based on self-report" the value of $P2FACTOR_{d_2}$was 1.270. $P2FACTOR_{d_2}$ was 1 for adults in all other Phase 2 domains.

The raked household-level weight was used as the foundation for calculating the replenishment sample adult weight. The adult Phase 1 base weight, denoted as $W4AP1BWT_{ijkl}$, was computed as the product of the three quantities shown in equation 5.1.4.3: the final household weight $W4HRKWT_{ijk}$; the reciprocal of the within-household probability of selection for adult l within Wave 4 household k of PSU i and segment j; and $P1FACTOR_{d_1}$ (calculated as shown in equation 5.4.1.1) for Phase 1 sampling domain d_1 of which adult l is a member:

$$W4AP1BWT_{ijkl} = (W4HRKWT_{ijk} \times \frac{1}{\text{Probability adult } l \text{ selected at Phase 1 from household } (ijk)}) \times P1FACTOR_{d_1} \quad (5.1.4.3)$$

The Wave 4 adult Phase 1 base weights were further adjusted for nonresponse to the Phase 2 screener. Weighting classes were formed using the same information for forming the nonresponse adjustment cells at the household level, in addition to the person-level data collected in the Wave 4 household screener as described in Section 5.1.1. The resulting adult weight (denoted as $W4AP1NRWT_{ijkl}$), adjusted for nonresponse between Phases 1 and 2 of the adult sampling procedure, was calculated for respondents to the Phase 2 screener as shown in equation 5.1.4.4:

$$W4AP1NRWT_{ijkl} = W4AP1BWT_{ijkl} \times W4AP1N_c \quad (5.1.4.4)$$

where $W4AP1N_c$ is the sum of $W4AP1BWT_{ijkl}$ for adults sampled at Phase 1 in weighting class c of which adult l is a member, divided by the sum of $W4AP1BWT_{ijkl}$ for all adults responding to the Phase 2 screener in that weighting class.

The Phase 2 weight, denoted as $W4AP2WT_{ijkl}$, was calculated as shown in equation 5.1.4.5:

$$W4AP2WT_{ijkl} = \text{W4}AP1NRWT_{ijkl} \times \frac{1}{\text{Probability adult } l \text{ from household } (ijk) \text{ selected at Phase 2}} \times P2FACTOR_{d_2}. \quad (5.1.4.5)$$

where $P2FACTOR_{d_2}$ is the Phase 2 probability adjustment factor for Phase 2 sampling domain d_2 of which adult l is a member as shown in equation (5.1.4.2).

Youth Weights

The number of youth respondents from the Wave 4 replenishment sample was also monitored to ensure that the targeted number of respondents would be met. This monitoring took into account not only the original targeted number for the replenishment sample but also the yield from the Wave 1 Cohort. If the yield from the Wave 1 Cohort was higher than expected, the sampling rate for the replenishment sample was adjusted to compensate.

At one point in the field period, the Wave 4 projected yield for youth was high enough that youth were not needed from every household in the replenishment sample. Partway through data collection, a subsampling rate was applied at the household level only for the purpose of selecting youth. This subsampling decision was subsequently reversed, i.e., all replenishment sample households screened later in the field period were considered eligible for youth selection, when projected yield from the Wave 1 Cohort decreased based on updated information. An adjustment factor was incorporated into the replenishment sample youth weighting process to account for the subsampling that occurred.

Similar to the adult weighting, the raked household-level weight was used as the foundation for calculating the replenishment sample youth weight. The youth base weight, denoted as $W4YBWT_{ijkl}$, was computed as the product of the three quantities as shown in equation 5.1.4.6: the final household weight $W4HRKWT_{ijk}$; the reciprocal of the probability that youth were eligible for selection from Wave 4 household k of PSU i and segment j , denoted as $W4PY_{ijk}$; and the reciprocal of the within-household probability of selection for youth l within household k of PSU i and segment j:

$$W4YBWT_{ijkl} = W4HRKWT_{ijk} \times \frac{1}{W4PY_{ijk}} \times \frac{1}{\text{Probability youth } l \text{ selected from household } (ijk)} \quad (5.1.4.6)$$

The youth base weights were adjusted to account for nonresponding youth. Weighting classes were formed using the same information for forming the nonresponse adjustment cells at the household

level, in addition to person-level data collected during the household screener. The nonresponse-adjusted weight for responding youth is shown in equation 5.1.4.7:

$$W4YNRWT_{ijkl} = W4YBWT_{ijkl} \times W4YN_c \quad (5.1.4.7)$$

where $W4YN_c$ is the nonresponse-adjustment factor calculated as the sum of $W4YBWT_{ijkl}$ for all sampled youth in weighting class c of which youth l is a member, divided by the sum of $W4YBWT_{ijkl}$ for all responding youth in that weighting class.

5.1.4.3.2 Compositing Weights

The weights of the respondents to the Wave 4 replenishment sample were combined with the weights of the Wave 4 respondents from the Wave 1 Cohort who were members of the CNP at Wave 4. Most members of the Wave 1 Cohort are also members of the Wave 4 Cohort, and most youth and adults selected for the replenishment sample at Wave 4 were eligible for selection at Wave 1. To account for persons having multiple chances of selection for the PATH Study, the sample weights were appropriately composited to reflect the Wave 4 CNP.

Addresses sampled at the time of Wave 1 were not given the opportunity to be selected for the Wave 4 replenishment sample. In addition, for addresses identified as having a chance of selection at Wave 1, their Wave 4 probabilities of selection for the replenishment sample were the same as for Wave 4 addresses that did not have a chance of selection at the time of Wave 1.

Questions were asked of the Wave 4 replenishment sample interview respondents to help determine if they were members of the U.S. CNP at the time of Wave 1. If so, they had two chances of selection for the PATH Study, through either the Wave 1 sample or the Wave 4 replenishment sample. To account for this, the weights of the Wave 4 Cohort members with two chances of selection were composited as discussed below. All other members of the Wave 4 Cohort had only a single chance of selection for the study (through the replenishment sample). See Section 5.1.4.3.3 for a description of the weights for the AYS replenishment sample interview respondents with a single chance of selection.

Establishing Initial Weights for the Compositing Process

For compositing, the weights used for the Wave 1 Cohort members who are also in the Wave 4 Cohort were the nonresponse-adjusted weights created as part of the single-wave weighting process described in Section 5.1.4.2. The weights used for the replenishment sample respondents who were also in the Wave 1 CNP were the nonresponse-adjusted weights described in Section 5.1.4.3.1. The two sets of weights were independently raked to the same population totals from the 2016 ACS PUMS. The raking was done using census region, age, race/ethnicity, sex, and educational attainment (for adults only).

The resulting preliminary raked weight, denoted as $W4PCRKWT_{ls}$, was calculated for person l from source s (Wave 1 Cohort or Wave 4 replenishment sample) as shown in equation 5.1.4.8:

$$W4PCRKWT_{ls} = \mathrm{W}4NRWT_{ls} \times W4PCR_{ls} \qquad (5.1.4.8)$$

where $W4PCR_{ls}$ is the preliminary raking adjustment factor for person l from source s. For the Wave 1 Cohort, $W4NRWT_{ls}$ is the nonresponse-adjusted weight created in the Wave 4 single-wave weighting process; for the Wave 4 Cohort, $W4NRWT_{ls}$ is $W4AP2WT_{ijkl}$ for adults and $W4YNRWT_{ijkl}$ for youth.

Creating and Applying Compositing Process

The compositing factors were calculated in two steps. First, effective sample sizes were calculated using the preliminary raked weights for each source (Wave 1 Cohort, Wave 4 replenishment sample) by respondent characteristics or "compositing domains." There were four compositing domains for adults, defined by the cross-classification of race (Black, non-Black) and age (ages 18-24, ages 25 and older), and six compositing domains for youth, corresponding to the single years of age from 12 to 17.[24]

Within a compositing domain m, the effective sample size for the Wave 1 Cohort, denoted as $n\_eff_m^1$, was calculated as the number of respondents (in the Wave 4 CNP) in domain m divided by

[24] The compositing domains for adults correspond to the dimensions of the adult sampling domains for which reliable population control totals were available for raking purposes.

the design effect associated with the variation of these respondents' preliminary raked weights, as shown in equation 5.1.4.9:

$$n\_eff_m^1 = \frac{n_m^1}{deff_m^1} \tag{5.1.4.9}$$

where n_m^1 is the number of respondents (in the Wave 4 CNP) from the Wave 1 Cohort in domain m, and $deff_m^1$ is the design effect associated with their preliminary raked weights, calculated as shown in equation 5.1.4.10:

$$deff_m^1 = 1 + (cv_m^1)^2 \tag{5.1.4.10}$$

where cv_m^1 is the coefficient of variation for the preliminary raked weights for the Wave 1 Cohort respondents (in the Wave 4 CNP) in domain m.

The effective sample size for a domain m of the Wave 4 replenishment sample, denoted as $n\_eff_m^4$, was calculated using the same formulae as described above, but based on the number of respondents from the replenishment sample (who were also in the Wave 1 CNP) in domain m and the design effect of their preliminary raked weights.

For a domain m of the Wave 1 Cohort, the compositing factor α_m $(0 < \alpha_m < 1)$ was then calculated as shown in equation 5.1.4.11:

$$\alpha_m = \frac{n\_eff_m^1}{n\_eff_m^1 + n\_eff_m^4}\,. \tag{5.1.4.11}$$

For the Wave 4 replenishment sample, the compositing factor for domain m was calculated as $(1 - \alpha_m)$. The values of α_m ranged from 0.720 to 0.895; the corresponding values of $(1 - \alpha_m)$ ranged from 0.280 to 0.105. These compositing factors were applied to the preliminary raked weights for the Wave 1 Cohort and the Wave 4 replenishment sample, respectively.

The composited weight, denoted as $W4CWT_{ls}$, was calculated for person l from source s (Wave 1 Cohort, Wave 4 replenishment sample) in domain m as the product of the preliminary raked weight and the compositing factor, α_m or $(1 - \alpha_m)$ for domain m, as shown in equation 5.1.4.12:

$$W4CWT_{lsm} = \begin{cases} W4PCRKWT_{lsm} \times \alpha_m, if\ s\ is\ Wave\ 1\ Cohort \\ W4PCRKWT_{lsm} \times (1 - \alpha_m), if\ s\ is\ replenishment\ sample \end{cases}. \tag{5.1.4.12}$$

5.1.4.3.3 Creating Final Wave 4 Weights for the Wave 4 Cohort

The Wave 4 Cohort final weights were created by first making some weight adjustments for the Wave 4 respondents not included in the compositing process and then combining them with the Wave 4 respondents with composited weights before the final raking and trimming step.

Weights for Wave 4 Cohort Members Not Included in the Compositing Process

The Wave 4 replenishment sample respondents who reported that they were not members of the Wave 1 CNP have a weight adjusted for nonresponse (as discussed in Section 5.1.4.3.1) but were not part of the compositing effort (discussed in Section 5.1.4.3.2) because they had only a single chance of being a member of the Wave 4 Cohort. With only a single chance of selection, the weights for this subgroup were generally larger than for other members of the Wave 4 Cohort whose weights were reduced by a compositing factor. There were two subgroups to which this set of respondents belonged: those in the military at Wave 1 and recent immigrants.

No issues with respect to the nonresponse-adjusted weights of those in the military at Wave 1 were identified. It was expected that recent immigrants would be concentrated in certain race/ethnicity groups. After examination of the distribution of the nonresponse-adjusted weights of recent immigrants by race/ethnicity, it appeared that they did not accurately represent recent immigrants at the time of Wave 4 in the population. So that these weights could more accurately represent this population, the weights were poststratified to corresponding 2016 ACS PUMS totals for persons who recently entered the U.S. The poststratification was performed separately for youth and adults. The control totals for youth were based on race/ethnicity; those for adults were based on the cross-classification of age group and race/ethnicity. After this step, a few outlier weights remained.

In order to ensure that such outliers would not have a disproportionately large impact on PATH Study estimates, a particular concern at the subgroup level, these poststratified non-composited weights were then evaluated in comparison to the distribution of the composited weights for the remainder of the Wave 4 Cohort. This was done separately for youth and adult respondents. Any non-composited youth weights that exceeded the largest composited youth weight among the other members of the Wave 4 Cohort were trimmed to the size of this largest youth weight. Corresponding trimming was performed for the poststratified non-composited adult weights. Fewer than 10 weights were trimmed in each group.

These cases (along with their poststratified and trimmed weights) were combined with the other Wave 4 Cohort adults and youth (along with their composited weights) prior to the last step of the weighting process, raking and trimming.

Raking and Trimming

The final raking and trimming step of the weighting process was conducted in an iterative fashion. First, the weights of all Wave 4 Cohort members were raked to independent population totals based on data from the 2016 ACS PUMS. For adults, the raking was done using cross-classifications of census region, age, race/ethnicity, sex, and educational attainment. For youth, the cross-classifications were based on census region, single-year of age, race/ethnicity, and sex. These variables were imputed if missing, using methods similar to those used at Wave 1. (See Section 5.3 for more information about this imputation.)

After raking, a trimming step was performed to bring any extreme weights down to the median weight plus four times the interquartile range within groups, or trimming cells. The trimming cells differed by recruitment wave, sampling domain, probability of selection (for those selected as youth or shadow youth), and whether the participant was in the Wave 1 CNP. After trimming, the sums of weights no longer matched the control totals, so raking and trimming were repeated until the resulting weights summed to the 2016 ACS PUMS totals for the raking dimensions and the weights were within the bounds defined by the interquartile range criterion.

After the iterative raking and trimming process, the final weight for person l, denoted as $W4CRKWT_l$, was calculated as shown in equation 5.1.4.13:

$$W4CRKWT_l = W4CWT_l \times W4RT_l \qquad (5.1.4.13)$$

where $W4RT_l$ is the combined raking and trimming adjustment for person l. The variable representing the final cross-sectional weight for a PATH Study participant in the Wave 4 Cohort responding to the Wave 4 adult interview ("A") or youth interview ("Y") is named R04_*x*_C04WGT. These variables are provided separated from the adult and youth/parent data files on weight files that also contain variables for variance estimation.

5.2 Variance Estimation

Accompanying the adult data files and youth/parent data files are replicate weights that may be used to estimate variances. The replication variance estimation approach is the preferred method for estimating variances from PATH Study data. The replicate weight variables for each wave, cohort, weight type, and interview type are summarized in Table 13 below.

Table 13. Replicate weight variables for each wave, cohort, weight type, and interview type*

Wave	Cohort	Weight type	Interview type	Variable names
1	Wave 1	Cross-sectional	Adult	R01_A_PWGT1 – R01_A_PWGT100
			Youth	R01_Y_PWGT1 – R01_Y_PWGT100
2	Wave 1	Single-wave/All-waves	Adult	R02_A_PWGT1 – R02_A_PWGT100
			Youth	R02_Y_PWGT1 – R02_Y_PWGT100
3	Wave 1	All-waves	Adult	R03_A_AWGT1 – R03_A_AWGT100
			Youth	R03_Y_AWGT1 – R03_Y_AWGT100
3	Wave 1	Single-wave	Adult	R03_A_SWGT1 – R03_A_SWGT100
			Youth	R03_Y_SWGT1 – R03_Y_SWGT100
4	Wave 1	All-waves	Adult	R04_A_A01WGT1 – R04_A_A01WGT100
			Youth	R04_Y_A01WGT1 – R04_Y_A01WGT100
4	Wave 1	Single-wave	Adult	R04_A_S01WGT1 – R04_A_S01WGT100
			Youth	R04_Y_S01WGT1 – R04_Y_S01WGT100
4	Wave 4	Cross-sectional	Adult	R04_A_C04WGT1 – R04_A_C04WGT100
			Youth	R04_Y_C04WGT1 – R04_Y_C04WGT100

*Replicate weight, pseudo-strata, and pseudo-PSU variables are included on the files with the questionnaire data for Wave 1 and Wave 2. For all other waves, these variables are on separate files corresponding to the respective wave, cohort, weight type, and interview type.

Replication variance estimation methods are increasingly used to provide analysts with a method for calculating standard errors of statistics. They produce consistent estimators of the variance for statistics that are smooth functions of estimated totals (Krewski and Rao, 1981); these include most commonly used statistics such as means, ratios, linear/logistic/Poisson regression coefficients, correlation coefficients, and many measures of association for categorical data. In most complex designs, such as the multi-stage sample design used in the PATH Study, the variance is estimated by assuming that the first-stage sampling is performed with replacement (Wolter, 2007).

The basic idea behind replication is to select subsamples repeatedly from the whole sample, calculate the statistic of interest for each subsample, and then use these subsamples or replicate statistics to estimate the variance of the full-sample statistic. Different ways of creating subsamples from the full sample result in different replication methods. The subsamples are called replicates and the statistics calculated from these replicates are called replicate estimates.

One major advantage of replication methods is that they can produce variance estimates for statistics that might not be available in standard software. Another advantage is that the replication variance estimation provides a simple way to account for adjustments that are made in weighting. As described above, the PATH Study full-sample weights are adjusted for nonresponse, trimming, and raking to control totals. By separately computing the weighting adjustments for each replicate, it is possible to reflect the effects of weight adjustments in the estimates of variance, which frequently results in a smaller variance because certain demographic estimates have been calibrated to known control totals (Valliant, 1993). Taylor series (linearization) methods for estimating the variance do not account for the variance reduction resulting from raking, and consequently often (but not always) provide variance estimates that are too large (Valliant, 2004; Chowdhury, 2013).

There are several ways of forming replicate weights, including balanced repeated replication (BRR), jackknife (JK-1, JK-2, and JK-n), and bootstrap. The choice of what kind of replicate weights to create is determined by the type of sampling design that was used to collect the data; in particular, whether or not stratification was used and how many PSUs were selected in each stratum. The replication method selected for the PATH Study is BRR (McCarthy, 1969).

The BRR method was selected for the PATH Study because (1) it allows calculation of the variance with fewer replicate weight variables than would be needed for the bootstrap for this dataset (this is possible because the subsets of PSUs for the replicates are carefully selected according to an orthogonal experimental design) and (2) BRR, unlike jackknife, produces consistent variance estimates for non-smooth statistics such as quantiles (Rao and Shao, 1993, 1999). The BRR method also allows correct estimation of standard errors for analyses involving quantiles and quantile regressions. This may be of interest for some of the biomarkers assessed by the PATH Study, such as the 10th percentile or the median of cotinine level. The PATH Study uses a variant of BRR known as Fay's method (Judkins, 1990). Fay's method produces more stability for the variance estimates for quantities in domains with small sample sizes. Using Fay's method, one half of the sample is weighted down by a factor ϵ (set to 0.3 for the PATH Study) and the remaining half is weighted up by a compensating factor $2 - \epsilon$.

Suppose that $\hat{\theta}$ is the full sample estimate of a population parameter θ, which can be a smooth or non-smooth function of a linear estimator. Then the variance estimator $v(\hat{\theta})$ takes the form as shown in equation 5.2.1 below:

$$v(\hat{\theta}) = c\sum_{g=1}^{G}(\hat{\theta}_{(g)} - \hat{\theta})^2 \quad (5.2.1)$$

where

$\hat{\theta}$ is the estimate of θ calculated using the full-sample weight,

$\hat{\theta}_{(g)}$ is the estimate of θ calculated using the *g*-th replicate weight,

G is the total number of replicates formed, and

c is a constant that depends on the replication method.

For standard BRR, $c = \frac{1}{G}$.

For Fay's method, $c = \frac{1}{G(1-\epsilon)^2}$, where ϵ is the factor used in Fay's method.

For the PATH Study, the value of the constant to be used when calculating the variance is

$$c = \frac{1}{100(1-0.3)^2} = 0.020408. \quad (5.2.2)$$

Creation of Variables for Variance Estimation

The first step in constructing the replicate weights was to create variables for pseudo-strata and pseudo-PSUs that reflect the variance structure. These were created using methods described in Korn and Graubard (2011, p. 206) and are called VARSTRAT and VARPSU on the data files.

There is a total of 92 strata and 156 PSUs in the design. Fifty of the strata have two PSUs sampled; these were left as is for the variance estimation. Seven strata have three PSUs sampled. Because the BRR method assumes that two PSUs are selected from each stratum, two pseudo-PSUs were created for each of these seven strata by randomly selecting two of the three PSUs in each stratum and combining them into one pseudo-PSU. The remaining PSU then became the second pseudo-PSU in the stratum.

Thirty-five of the strata have one PSU that is selected with certainty and called self-representing (SR). In these strata, the variability comes from the secondary sampling units (segments), which were selected from a list of segments ordered such that similar segments were close together (see Section 2.1.2.2). The following procedure was used to create pseudo-strata and pseudo-PSUs within the SR PSUs for variance estimation purposes.

1. Each SR PSU with fewer than 60 segments was treated as one pseudo-stratum with two pseudo-PSUs. The segments in the SR PSU were assigned to the two pseudo-PSUs so that the odd-numbered segments in the sorted list were assigned to one pseudo-PSU and the even-numbered segments in the sorted list were assigned to the other pseudo-PSU. In this way, adjacent segments in the sorted list were assigned to different pseudo-PSUs.

2. Four of the SR PSUs had large numbers of segments selected, and these were divided into pseudo-strata for variance estimation purposes. The pseudo-strata were formed by cutting the ordered list of segments into the desired number of pseudo-strata. Then, the segments within the pseudo-strata were assigned to pseudo-PSUs as described in (1).

Applying the BRR method to the PATH Study's pseudo-strata and the initial household weights yields 100 initial replicate weights. These 100 initial replicate weights were adjusted using the steps described in Section 5.1 for the full-sample weights to arrive at the final set of corresponding replicate weights for variance estimation.

Software Options

The data files are provided in several formats. When the data analysis software package allows, the replicate weights should be used for all variance calculations to reflect the impact of the complex sample design and the various weighting adjustments on standard errors. Note that, although variables for pseudo-strata (VARSTRAT) and pseudo-PSUs (VARPSU) are included with the data files, variance estimates calculated using these variables with linearization (for example, by using the STRATA and CLUSTER statements in SAS®) do not reflect the impact of the weighting adjustments and may result in incorrect inferences. Some example SAS, SUDAAN®, Stata®, R, and SPSS® program code for generating popular statistics is provided in Appendix A.

The BRR replication method of variance estimation is available in both the SAS and Stata software packages. SPSS Complex Samples™ currently does not offer BRR or other forms of replication-based estimation. Note that the open-source R statistical software language does not have core programs that do survey data analysis. There are various contributed packages to R (such as the survey package) that analyze data from complex surveys and handle replication methods for variance estimation; however, these contributed packages are not peer-reviewed or subject to quality standards, and their features are subject to change in future versions of R.

Although software packages do not universally accommodate replicate weights for all analytic methods, the replication method can be applied by repetition to any analytic routine. That is, the desired analysis would first be run using the full-sample weight. Then, it would be repeated replacing the full-sample weight by each replicate weight, in turn (i.e., 100 times for PATH). The formula for BRR variance estimates (provided in equation 5.2.1) would then be used to estimate the variance of any parameter (e.g., regression coefficient) of interest.

Variance estimates for small domains may be unstable because some PSUs may contain no observations belonging to the domain. For small domains, the variance estimates produced by statistical software packages may differ because they use different methods to adjust for strata in which only one PSU contains domain members. For more information, see Graubard and Korn (1996), Lohr (2010, p. 570), or Lewis (2013).

5.3 Imputation

Demographic characteristics of adults and youth selected for the PATH Study were used in the creation of the weights. These included variables indicating sex, age, education level (for adults only), race, and ethnicity of the sampled persons. However, because some of this information may be missing for some sampled adults and youth, imputation methods were used to assign values when self-reported information was not available.

5.3.1 Wave 1 Imputation

For both adults and youth, imputation was performed by first considering information provided in the household screener and then by using statistical imputation methods. The imputation methods were performed for respondents and nonrespondents[25] because data for both were needed for weighting, but only the values for respondents are included on the data files. There was no imputation performed for Waves 2 and 3 because the weighting process relied on demographic and other characteristics from Wave 1.

The sections below provide the methods used for imputing sex (Section 5.3.1.1), age (Section 5.3.1.2), education (Section 5.3.1.3), race (Section 5.3.1.4), and ethnicity (Section 5.3.1.5). Section 6.4

[25] These are adults and youth belonging to responding households who did not respond to the interview.

provides a description of the variable naming convention used for the adult data files and youth/parent data files.

5.3.1.1 Sex

Sex was assigned for adults and youth from information provided in the extended interview (questionnaire item R01_AM0004 for adults and item R01_YM0004 for youth). If this was not available, either because the sampled person refused to provide it or because the sampled person did not respond to the interview, sex was assigned based on the information provided in the household screener.

After reviewing these sources, however, sex remained unavailable for a small number of sampled persons. Common demographic variables (for example, census region, age, education, race) are not indicators of sex. In most cases, this information was also unavailable if information on sex could not be obtained from the extended interview or household screener, so sex was randomly assigned so that approximately one half of the cases with missing sex was assigned as female and the other half was assigned as male.

The values of R01_AM0004 and R01_YM0004 are provided in the variables R01R_A_SEX for adults and R01R_Y_SEX for youth. All values, imputed and unimputed, are contained in the variable R01R_*x*_SEX_IMP, where *x* is "A" (adult) or "Y" (youth); the variable R01R_*x*_SEX_IMPFLAG indicates which values are from the interview, i.e., not imputed (R01R_*x*_SEX_IMPFLAG = 0), from the household screener (R01R_*x*_SEX_IMPFLAG = 1), or the result of the random imputation method described above (R01R_*x*_SEX_IMPFLAG = 2).

5.3.1.2 Age

A single-year of age variable was created for both adults and youth. For adults, age was first calculated based on the date of birth provided during the interview. If date of birth was not provided in the interview, the age in years was used if available.[26] For youth, age was first calculated

[26] Only adults who did not provide their date of birth were asked to provide their age in years.

based on the date of birth provided by the parent during the consent process. If date of birth was not provided, the age in years (also requested during the consent process) was used.

If age was still missing (because the responding adult refused to provide their date of birth or age, the responding youth was an emancipated minor and no parental consent was required, or the sampled person did not respond to the extended interview), age in years provided in the household screener was used.

For nonresponding youth, if age was missing after the above processing, a value was statistically imputed since age-in-years was used in the youth weighting process. Given that the only information available about these youth is from the household screener and that information (such as race, ethnicity, or sex) is not indicative of a child's age, the age in years was randomly assigned using the youth population distribution according to the 2013 ACS PUMS.

For adults, a four-level age category (18-24, 25-44, 45-64, and 65 and older) was used for the weighting process, but was not included on the data file. If the imputed single year of age was missing after the above processing, but the sampled adult indicated in the Phase 2 screener that they were in one of the age categories under 30 years old, that age category was used.[27] If that information was not available, the age category indicated in the household screener was used.

As with youth, if none of this information was available for an adult, the age category was imputed using the population distribution according to the 2013 ACS PUMS data. Because a higher proportion of adults in the older age categories are female, the random assignment was conducted separately by sex. Age range information, provided in the Phase 2 screener, was used for some adults who indicated they were 30 years or older. For these adults, the age category was randomly assigned using the population distribution by sex according to the 2013 ACS PUMS data for that age group. If it was unknown whether the sampled adult was 30 or older, the age category was randomly assigned using the population distribution by sex for those 18 and older.

Neither the single year of age variable nor its imputed counterpart is provided on the PUFs. Instead, age range variables, R01R_A_AGECAT7 for adults and R01R_Y_AGECAT2 for youth, are

[27] Questionnaire item R01_AM0003 requesting an age category was asked in the interview only if the sampled adult refused to provide a date of birth or age in years. The highest age category was "30 or older."

provided along with their imputed counterparts R01R_A_AGECAT7_IMP and R01R_Y_AGECAT2_IMP, respectively. The age categories were created from the single-year of age variables, with the exception of adults with missing single-year of age who indicated that they were in one of the age categories under 30 years old in the Phase 2 screener; these cases were assigned to the appropriate age category based on that information as described above. All values for respondents, whether imputed or not, are contained in the imputed variables; the variable R01R_*x*_AGECAT#_IMPFLAG indicates which values are from the interview (R01R_*x*_AGECAT#_IMPFLAG = 0), from the household screener (R01R_*x*_AGECAT#_IMPFLAG = 1), or are missing because no further information is available (adults only) (R01R_*x*_AGECAT#_IMPFLAG = 3), where *x* is "A" (adult) or "Y" (youth) and # is "2" or "7" as appropriate.

5.3.1.3 Education (Adults Only)

Adult respondents were asked questionnaire item R01_AM0018 regarding the highest level of education they attained. A five-level education variable was used for the adult weighting and created by collapsing the 11 response categories as follows: less than high school or GED; high school graduate; some college but no degree, or associates degree; bachelor's degree; advanced degree. Missing values were imputed using hot deck imputation because demographic characteristics that are indicative of educational attainment were available.[28] The imputation cells were formed by cross-classifying categories of census region, age, and sex, using imputed values of age and sex (calculated as described in the preceding sections) as appropriate. Neither the education variable described above nor its imputed counterpart is provided on the adult PUF. Instead, a six-level education variable (R01R_A_AM0018) is provided with a separate category for respondents indicating "GED" as their highest level of education.

5.3.1.4 Race

A four-level race variable was used for both the adult and youth weighting. In the extended interview, respondents were asked to indicate which of 14 race categories applied to them (questionnaire items with the prefix R01_AM0006 for adults and items with the prefix R01_YM0006

[28] Hot deck imputation is a method for handling missing data in which each missing value is replaced with an observed response from a "similar" unit in the imputation cell.

for youth). The responses were combined into a single variable with four categories: white alone, Black or African American alone, Asian alone (including multiple Asian categories), and other (including multi-racial).

If the sampled adult or youth did not respond to the race question, or did not respond to the extended interview, the race information was assigned according to the information provided in the household screener.[29] The race information gathered for the sampled person in the household screener was considered first; if that was not available, the race information that the household screener respondent reported about himself or herself was used as a proxy for the sampled person.

If none of this information was available, race was imputed using hot deck imputation because demographic characteristics that are indicative of race were available. For adults, imputation cells were formed by cross-classifying categories of census region, age, and education, using imputed values of age and education (calculated as described in the preceding sections) as appropriate; for youth, the hot deck imputation cells were formed by the four categories of census region.

Variables indicating three levels of race (white alone, Black alone, and other) are provided on the PUFs. All values, imputed and unimputed, are contained in the variable R01R_*x*_ RACECAT3_IMP, where *x* is "A" (adult) or "Y" (youth). The variables R01R_A_RACECAT3_IMPFLAG and R01R_Y_RACECAT3_IMPFLAG indicate which values are from the interview (R01R_*x*_RACECAT3_IMPFLAG = 0), from the household screener (R01R_*x*_RACECAT3_IMPFLAG = 1), or the result of the imputation method described above (R01R_*x*_RACECAT3_IMPFLAG = 2).

5.3.1.5 Ethnicity

A two-level ethnicity variable indicating whether the sampled person is of Hispanic origin was used for both the adult and youth weighting. This variable was initially created from the Hispanic origin question asked in the extended interview (questionnaire item R01_AM0005_01 for adults and item R01_YM0005_01 for youth). The variables R01R_A_HISP and R01R_Y_HISP indicate whether the respondent is of Hispanic origin based on these variables for adults and youth, respectively.

[29] There were only five race category responses in the household screener, but it was possible to uniquely code them into the four race categories used in weighting.

If the sampled adult or youth did not respond to this question, or did not respond to the extended interview, the ethnicity information was assigned according to the information provided in the household screener: The ethnicity information gathered for the sampled person in the household screener was considered first; if that was not available, the ethnicity information that the household screener respondent reported about himself or herself was used as a proxy for the sampled person.

If none of this information was available, ethnicity was imputed using hot deck imputation because demographic characteristics that are indicative of ethnicity were available. For adults, imputation cells were formed by cross-classifying categories of census region, age, and education, using imputed values of age and education (calculated as described in the preceding sections) as appropriate. In Wave 1, while the ethnicity variable was imputed for some adults, imputation was not necessary for the youth.

All values for respondents, imputed or not, are contained in the variable R01R_*x*_HISP_IMP, where *x* is "A" (adult) or "Y" (youth). The variables R01R_A_HISP_IMPFLAG and R01R_Y_HISP_IMPFLAG indicate which values are from the interview (R01R_*x*_HISP_IMPFLAG = 0), from the household screener (R01R_*x*_HISP_IMPFLAG = 1), or the result of the imputation method described above (R01R_*x*_HISP_IMPFLAG = 2).

5.3.2 Wave 4 Imputation

Respondents to Wave 4 of the PATH Study were from two separate samples selected at different times: from the replenishment sample and the Wave 1 Cohort. The imputation approach differed for these two groups. For respondents from the replenishment sample, imputation was performed by first considering information provided in the Wave 4 household screener and then by using statistical imputation methods. For respondents from the Wave 1 Cohort, imputed values from Wave 1 were used in addition to new statistical imputation after considering information provided in previous waves. Even though Wave 1 Cohort members who are not in the Wave 4 Cohort were excluded from the Wave 4 Cohort cross-sectional weighting process, they were included in the imputation process so that the imputed variables could be made available for all Wave 4 respondents.

Non-missing values for sex, age, highest education, race, and ethnicity were needed for raking the Wave 4 Cohort cross-sectional weights to control totals. Imputation was performed only for Wave 4 respondents because nonrespondents are not included in the raking procedure.

All Wave 4 respondents had non-missing values for age. The sections below provide the methods used for imputing sex (Section 5.3.2.1), education (Section 5.3.2.2), race (Section 5.3.2.3), and ethnicity (Section 5.3.2.4). The imputation of these Wave 4 characteristics was performed in the order they are presented below except that for the Wave 1 Cohort, race and ethnicity were imputed prior to education. This sequential approach allowed the imputed values for earlier variables to be used in the imputation procedure for later variables, thus preserving correlations among the characteristics.

5.3.2.1 Sex

The variables R04R_A_SEX and R04R_Y_SEX contain the sex responses for adults and youth, respectively. All values, imputed and unimputed, are contained in the variable R04R_*x*_SEX_IMP, where *x* is "A" (adult) or "Y" (youth); the variable R04R_*x*_SEX_IMPFLAG indicates which values are from the interview, i.e., not imputed (R04R_*x*_SEX_IMPFLAG = 0), from the household screener (R04R_*x*_SEX_IMPFLAG = 1), or the result of the imputation method described below (R04R_*x*_SEX_IMPFLAG = 2). The sections below describe the imputation of missing values of sex for the replenishment sample and the Wave 1 Cohort, respectively.

Replenishment Sample

Sex was assigned for adults and youth by first looking at the information provided in the Wave 4 extended interview (questionnaire item R04_AM0004_RS for adults and item R04_YM0004_NB for youth). If this information was not provided, then sex was assigned according to the information provided in the Wave 4 household screener.

Wave 1 Cohort

Members of the Wave 1 Cohort were asked to self-report their sex at Wave 1 only if they were an adult or youth at that time. This information was not requested from those who were shadow youth at Wave 1 until they completed their first youth interview at a subsequent wave. For Wave 4 adults,

sex was assigned by first considering information provided in the Wave 1 extended interview (questionnaire item R01_AM0004 for adults and item R01_YM0004 for youth). If sex was still missing, then the Wave 1 imputed value of sex was assigned. For Wave 4 youth, sex was assigned by initially considering information provided in their first youth interview (R01_YM0004 for Wave 1, R02_YM0004_NB for Wave 2, R03_YM0004_NB for Wave 3, and R04_YM0004_NB for Wave 4). If sex was still missing, then the Wave 1 imputed value of sex was assigned.

5.3.2.2 Education (Adults Only)

The variable R04R_A_EDUC4 contains the four education categories used in weighting before imputation. All values, imputed and unimputed, are contained in the variable R04R_A_EDUC4_IMP. The variable R04R_A_EDUC4_IMPFLAG indicates which values are from the interview (R04R_A_EDUC4_IMPFLAG = 0) or the result of the imputation method described below (R04R_ A_EDUC4_IMPFLAG = 2). The sections below describe the imputation of missing values of education for the replenishment sample and the Wave 1 Cohort, respectively.

Replenishment Sample

All Wave 4 adult respondents were asked for their highest level of education, with 11 valid response categories presented.[30] Responses were combined into four categories for weighting purposes: less than high school or GED; high school graduate; some college but no degree, or associates degree; bachelor's degree and beyond. Missing education values were imputed using hot deck imputation because demographic characteristics that are indicative of educational attainment were available. Imputation cells were formed by cross-classifying categories of census region, age, and sex, using imputed values of sex (calculated as described in the preceding section) as appropriate.

Wave 1 Cohort

The four-level education variable described above was also created for adult respondents from the Wave 1 Cohort, based on the same questionnaire item responses. Missing education values were imputed using hot deck imputation. However, because the educational attainment item was also

[30] If the adult completed a parent interview prior to their adult interview, this information was captured in questionnaire item R04_PM0001; otherwise, this information was captured in questionnaire item R04_AM0018.

included in the Wave 1, 2, and 3 adult interviews, available historical information was used in the imputation process. The imputation cells were formed by cross-classifying categories of age, sex, race/ethnicity,[31] imputed highest education level at Wave 1, highest education level at Wave 2, and highest education level at Wave 3, using imputed values of sex and race/ethnicity (calculated as described in the preceding sections) as appropriate. As a final check, any imputed value less than the highest education level reported in the previous wave was replaced with the previous wave's value.

5.3.2.3 Race

The variables R04R_A_RACECAT3 and R04R_Y_RACECAT3 contain the race responses for adults and youth, respectively. All values, imputed and unimputed, are contained in the variable R04R_*x*_RACECAT3_IMP, where *x* is "A" (adult) or "Y" (youth). The variables R04R_A_RACECAT3_IMPFLAG and R04R_Y_RACECAT3_IMPFLAG indicate which values are from the interview (R04R_*x*_RACECAT3_IMPFLAG = 0), from the household screener (R04R_*x*_RACECAT3_IMPFLAG = 1), or the result of the imputation method described below (R04R_*x*_RACECAT3_IMPFLAG = 2). The sections below describe the imputation of missing values of race for the replenishment sample and the Wave 1 Cohort, respectively.

Replenishment Sample

In the extended interview, respondents were asked to select as many of the 14 valid race categories as applied to them (questionnaire items with the prefix R04_AM0006_RS for adults and items with the prefix R04_YM0006_NB for youth). Responses were combined into a single variable with four categories (White alone; Black alone; Asian alone, including multiple Asian categories; other, including multi-racial) for weighting purposes.

If the responding adult or youth did not answer the race question, race was assigned according to the information provided in the Wave 4 household screener.[32] The race information gathered for the respondent in the household screener was considered first; if that was not available, the race

[31] For the Wave 1 Cohort, race and ethnicity were imputed prior to the imputation of education.

[32] There were five race categories presented in the household screener but they may be uniquely coded into the four race categories used in weighting.

information that the household screener respondent reported about himself or herself was used as a proxy for the respondent's race.

If none of this information was available, race was imputed using hot deck imputation because demographic characteristics that are indicative of race were available. For adults, imputation cells were formed by cross-classifying categories of census region, age, and education, using imputed values of education (calculated as described in the preceding section) as appropriate; for youth, imputation cells were formed by the four categories of census region.

Wave 1 Cohort

Members of the Wave 1 Cohort were asked to self-report their race at Wave 1 only if they were an adult or youth at that time. This information was not requested from those who were shadow youth at Wave 1 until they completed their first youth interview at a subsequent wave. For Wave 4 adults, race was assigned by first considering information provided in the Wave 1 extended interview (questionnaire items with the prefix R01_AM0006 for adults and items with the prefix R01_YM0006 for youth). If race was still missing, then the Wave 1 imputed value of race was assigned. For Wave 4 youth, race was assigned by initially considering information provided in their first youth interview (items with prefix R01_YM0006 for Wave 1, items with prefix R02_YM0006_NB for Wave 2, items with prefix R03_YM0006_NB for Wave 3, and items with prefix R04_YM0006_NB for Wave 4). If race was still missing, then the Wave 1 imputed value of race was assigned.

5.3.2.4 Ethnicity

The variables R04R_A_HISP and R04R_Y_HISP contain the ethnicity responses for adults and youth, respectively. All values for respondents, imputed or not, are contained in the variable R04R_*x*_HISP_IMP, where *x* is "A" (adult) or "Y" (youth). The variables R04R_A_HISP_IMPFLAG and R04R_Y_HISP_IMPFLAG indicate which values are from the interview (R04R_*x*_HISP_IMPFLAG = 0), from the household screener (R04R_*x*_HISP_IMPFLAG = 1), or the result of the imputation method described below (R04R_*x*_HISP_IMPFLAG = 2). The sections below describe the imputation of missing values of ethnicity for the replenishment sample and the Wave 1 Cohort, respectively.

Replenishment Sample

In the extended interview, respondents were asked to select as many of the five valid ethnicity categories as applied to them (questionnaire items with the prefix R04_AM0005_RS for adults and items with the prefix R04_YM0005_NB for youth). Responses were combined into a single variable with two categories (Hispanic, not Hispanic) for weighting purposes.

If the responding adult or youth did not answer the ethnicity question, ethnicity was assigned according to the information provided in the Wave 4 household screener. The ethnicity information gathered for the respondent in the household screener was considered first; if that was not available, the ethnicity information that the household screener respondent reported about himself or herself was used as a proxy for the respondent's ethnicity.

Wave 1 Cohort

Members of the Wave 1 Cohort were asked to self-report their ethnicity at Wave 1 only if they were an adult or youth at that time. This information was not requested from those who were shadow youth at Wave 1 until they completed their first youth interview at a subsequent wave. For Wave 4 adults, ethnicity was assigned by first considering information provided in the Wave 1 extended interview (questionnaire items with the prefix R01_AM0005 for adults and items with the prefix R01_YM0005 for youth). If ethnicity was still missing, then the Wave 1 imputed value of ethnicity was assigned. For Wave 4 youth, ethnicity was assigned by initially considering information provided in their first youth interview (items with prefix R01_YM0005 for Wave 1, items with prefix R02_YM0005_NB for Wave 2, items with prefix R03_YM0005_NB for Wave 3, and items with prefix R04_YM0005_NB for Wave 4). If ethnicity was still missing, then the Wave 1 imputed value of ethnicity was assigned.

5.4 Selecting the Appropriate Weights

This section provides information to assist users in determining appropriate approaches for the analysis of the PATH Study data and the weights to use with each approach. For example, estimates can be made comparing tobacco use reported in the latest wave to that used in Wave 1 or comparing all prior waves for the same set of individuals. These are longitudinal comparisons, and the weights used would be associated with the latest wave.

As another example, estimates can be obtained for the CNP at Wave 4, and these can be compared to the same estimates representing the CNP at a prior wave, based on those who responded at that wave. This is a cross-sectional comparison and some participants may have responded in one wave but not the other. The PATH Study was not designed with a focus on cross-sectional estimation and comparisons. However, if variance calculations are correctly specified, they will correctly reflect the correlation between the overlapping sets of respondents and the estimates will be appropriate but approximate. Cross-sectional comparisons between waves should be undertaken with care, and are not appropriate in all situations. Please refer to Section 5.4.3.3 for more information.

Section 5.4.1 considers different designs that observational studies may use to measure change over time. Sections 5.4.2 and 5.4.3 present the target populations of inference and analytic considerations for the PATH Study. Regardless of the analysis, it is important that analysts use the appropriate framework and methods of estimating standard errors to appropriately reflect the complex sample design employed for the PATH Study. Please refer to Appendix A for example program code illustrating the correct specifications for creating appropriate variance estimates that reflect this complex sample design.

5.4.1 Background

The PATH Study is a longitudinal cohort study, not a repeated cross-sectional study. In a repeated cross-sectional study, such as the National Health Interview Survey,[33] a sample of persons is selected at time 1, and that sample is used to produce estimates for variables of interest at time 1. For example, an estimate of interest might be the percentage of 18-year-olds who use e-cigarettes. Then, at time 2, a new sample is selected, and the percentage of 18-year-olds using e-cigarettes is estimated from that new sample. The sampling may be repeated as many times as desired. Each sample provides a snapshot of what is happening in the target population or subpopulation at the time the sample is taken, but it does not provide information about a particular sample at other times. Change is estimated by comparing the time-1 sample with the time-2 sample. Thus, repeated cross-sectional samples may be able to provide information on whether e-cigarette usage among the subpopulation of 18-year-olds increased or decreased at time 2 compared to time 1. However, they

[33] U.S. DHHS, CDC/National Center for Health Statistics.

do not permit analysis of change for the same set of persons over time, for instance, as to whether 18-year-old e-cigarette users at time 1 are still using the product at time 2.

In a longitudinal cohort study, such as the PATH Study, a sample of persons is selected at time 1 (just as in a repeated cross-sectional study). At time 2 and subsequent points in time, however, the same persons are re-interviewed. They may be asked the same questions as at time 1, and they may also be asked different questions at later interviews. Thus, a longitudinal study provides information on how individual participants in the study change over time, and allows the researcher to investigate the individual trajectories experienced by different people over time. A longitudinal study allows analysis of "within-person" change over time and can help researchers understand persistence of behavior or factors associated with changing behaviors and outcomes.

A longitudinal study can provide important precision gains for assessing change when compared with a repeated cross-sectional study. Assessing change over time "within person" takes maximum advantage of the correlation between measures on the same respondents over time to increase precision and power. An issue with a longitudinal study is that it does not include new members of the population such as recent immigrants. Rather, it includes only persons who were eligible for the study at the time of recruitment. To help address this issue, the PATH Study design plan includes periodic "refreshment" of the sample, randomly selecting new study participants from the U.S. CNP. The first refreshment (or replenishment) sample was selected at Wave 4. The second replenishment sample will be selected at Wave 6.

5.4.2 Target Populations

The target population for a study is the population for which the sample weights can be used for inferential purposes. The sample that represents the target population for a longitudinal study is generally referred to as a cohort and the target population for that sample at a given point in time is the population from which the sample was selected that remains eligible at the time in question. The study currently follows two cohorts, the Wave 1 Cohort and the Wave 4 Cohort. Descriptions of these cohorts and their target populations at each wave are in Chapter 2.

Generally, for the PATH Study, the target population for a given cohort is the CNP of a particular age range at the time of sampling. For the waves that follow, the target population is the same

population who at that wave still resides in the U.S. and is not incarcerated. However, the participants at later waves do not represent some members of the CNP at the time of that wave. For example, at Wave 2, the Wave 1 Cohort does not represent immigrants who were not residents at Wave 1 or persons who were in the military or an institution (e.g., health care or in jail) at the time of Wave 1 who re-joined the U.S. CNP at Wave 2. This means that in follow-up waves, the PATH Study cohorts do not perfectly align with the CNP at the time of follow-up.

However, the subpopulations that form this "over" or "under" coverage are relatively small for follow-up waves close in time to the formation of the cohort. So, the weights developed for the cohort for some follow-up waves are suitable for producing useful cross-sectional estimates of the CNP at the time of the respective wave. For example, weights developed for Wave 2 and Wave 3 are suitable for cross-sectional estimation for those waves.

With the establishment of each new cohort, weights that truly represent the CNP at that wave are developed so those weights should be used for cross-sectional estimation for that wave and for subsequent waves until another new cohort is established. For example, the Wave 4 weights for the Wave 4 Cohort should be used for cross-sectional estimation for the time of Wave 4. Because weights developed for the Wave 4 Cohort will more closely represent the population at the time of subsequent waves than weights for the Wave 1 Cohort, weights developed for the Wave 4 Cohort should be used for cross-sectional estimation for follow-up waves until a new cohort is established at Wave 6.

5.4.3 Analytic Considerations

As the PATH Study is a longitudinal study, there will be many analyses examining within-person transitions over time with respect to the use of tobacco products, and changes in health, tobacco-related attitudes, and other factors. Thus, for most analyses, the focus will be on the characteristics of the respondents in the latest wave, how those may have changed from previous waves, and what factors may have been associated with such changes. For each wave after Wave 1, longitudinal weights are available to permit such analyses. These types of analyses are discussed more in Section 5.4.3.1 and 5.4.3.2.

Cross-sectional weights are available for estimation of the Wave 1 and Wave 4 target populations. As discussed in Section 5.4.2, some weights, although longitudinal in nature, can be used to approximate the cross-sectional population for the respective wave. These types of analyses, and when they are appropriate, are discussed more in Sections 5.4.3.3 and 5.4.3.4.

For the PATH Study analyses, the recommended approach for computing sampling errors is to employ the replicate weights provided for each weight type. If an analyst chooses to use Taylor Series methods and estimates are being made for subdomains of interest, then all records in the full sample should be retained in the analysis data file and a domain analysis conducted so that the software package employed can appropriately compute variance estimates.

Table 14 summarizes the weights available for each wave and cohort, for different analysis types, including which PATH Study participants were assigned each weight. (Note that if there is only one set of longitudinal weights, such as with Wave 2, that set of weights serves as all-waves and single-wave weights.) The appropriate weight for any analysis depends on the PATH Study participants included in the analysis (e.g., Wave 2 adult interview respondents), and not any particular analysis tool (e.g., logistic regression).

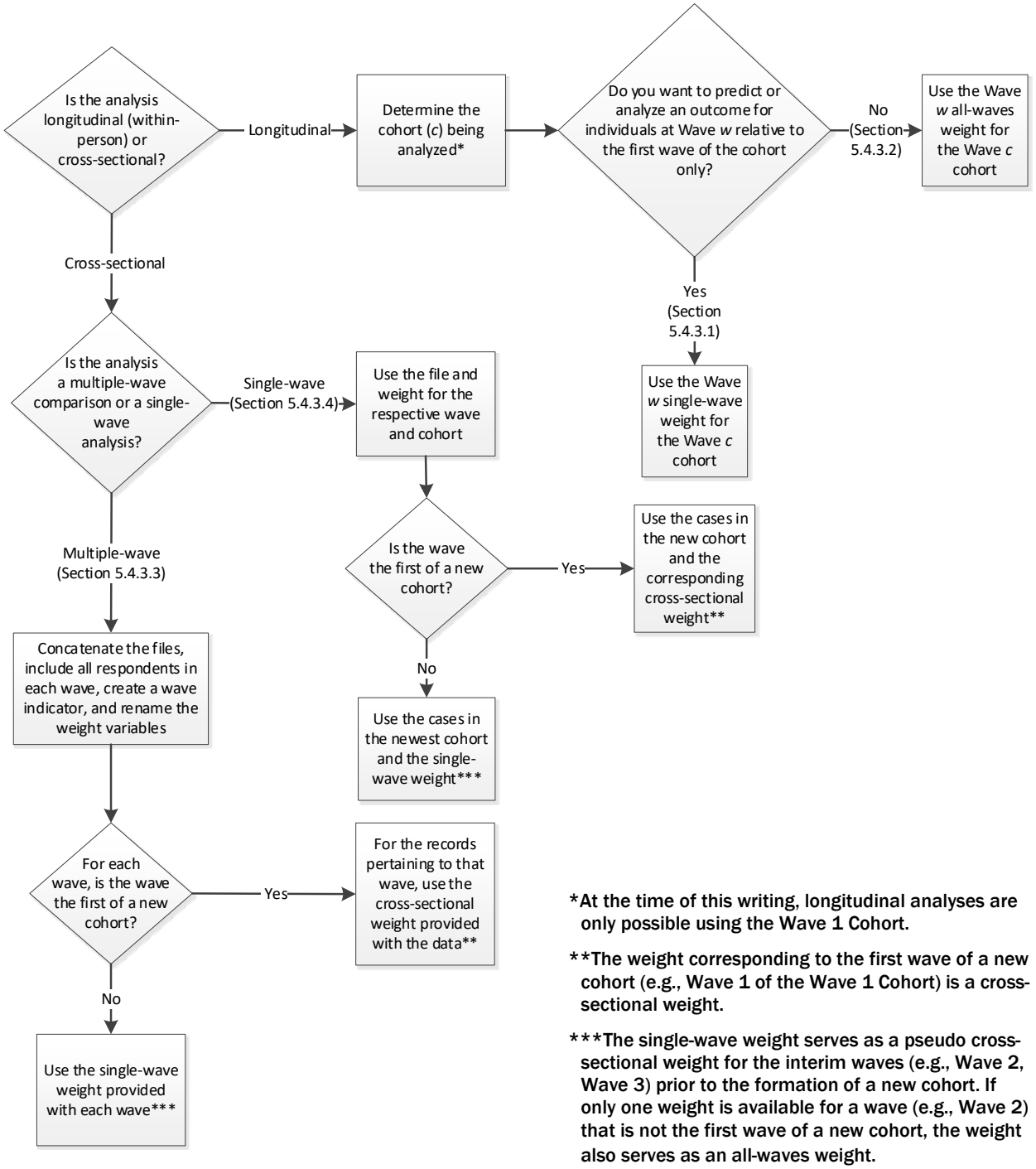
Table 14. Descriptions of weights available for each wave and cohort

Wave	Cohort	Weight type	Analysis type	Assigned to…
1	Wave 1	Cross-sectional*	Cross-sectional	Wave 1 interview respondents in the Wave 1 Cohort
2	Wave 1	Single-wave/ All-waves	Longitudinal, cross-sectional**	Wave 2 interview respondents in the Wave 1 Cohort
3	Wave 1	All-waves	Longitudinal	Wave 3 interview respondents in the Wave 1 Cohort who also participated in Wave 2
3	Wave 1	Single-wave	Longitudinal, cross-sectional**	Wave 3 interview respondents in the Wave 1 Cohort
4	Wave 1	All-waves	Longitudinal	Wave 4 interview respondents in the Wave 1 Cohort who also participated in Wave 2 and Wave 3
4	Wave 1	Single-wave*	Longitudinal	Wave 4 interview respondents in the Wave 1 Cohort who completed a Wave 1 interview
4	Wave 4	Cross-sectional	Cross-sectional	Wave 4 interview respondents in the Wave 4 Cohort; no youth selected in the shadow sample at Wave 4 have this weight

* No youth selected in the shadow sample at Wave 1 have this weight.
** The single-wave weight serves as a pseudo cross-sectional weight for the interim waves (e.g., Wave 2, Wave 3) prior to the formation of a new cohort (see Section 5.4.2).

The flowchart in Figure 2 summarizes the types of decisions researchers will make in analyzing PATH Study data. This is further discussed, with examples of analyses of each type, in the sections below.

Figure 2. Flowchart providing guidance on which weights to use for analyses of PATH Study data



***At the time of this writing, longitudinal analyses are only possible using the Wave 1 Cohort.**

****The weight corresponding to the first wave of a new cohort (e.g., Wave 1 of the Wave 1 Cohort) is a cross-sectional weight.**

*****The single-wave weight serves as a pseudo cross-sectional weight for the interim waves (e.g., Wave 2, Wave 3) prior to the formation of a new cohort. If only one weight is available for a wave (e.g., Wave 2) that is not the first wave of a new cohort, the weight also serves as an all-waves weight.**

5.4.3.1 Longitudinal Analyses Using Data From the First Wave of a Cohort in Estimating a Later Wave Outcome

This section provides examples where the desired analysis is to relate a behavior or outcome at a later wave (Wave *w*) to characteristics of the same person at the first wave of a cohort (i.e., Wave 1 for the Wave 1 Cohort). In this instance, the later wave's single-wave weight is to be used.

The defining characteristic for this type of analysis is that the interest is in the trajectory of the same individuals traced from the first wave of the cohort to Wave *w*. The goal is to estimate at Wave *w* the extent to which individuals have changed since the first wave of the cohort based on characteristics from that wave and Wave *w* only. These are within-person analyses for which the PATH Study was designed.

Because no data are collected for shadow youth at their recruitment wave, Wave *c*, they may not be included in such an analysis of Wave *c* cohort data. For example, Wave 1 Cohort participants selected as shadow youth at Wave 1, cannot be included in an analysis of data using characteristics from Wave 1 and Wave *w* because they have no Wave 1 data available for analysis.

Examples of analyses using Wave 1 information to estimate a Wave *w* outcome for the Wave 1 Cohort:

- Estimating the proportion of Wave *w* e-cigarette users who used e-cigarettes at Wave 1. This is a longitudinal analysis using the persons who have data at both Wave 1 and Wave *w*.
- Calculating an odds ratio (a ratio comparing (the odds of being an e-cigarette user at Wave *w* if the person was an e-cigarette user at Wave 1) to (the odds of being an e-cigarette user at Wave *w* if the person did not use e-cigarettes at Wave 1)). The target population is that at Wave *w*. The odds ratio can be calculated from estimated proportions obtained from the two by two contingency table below:

E-cigarette use status	Nonuser at Wave *w*	User at Wave *w*	Total
Nonuser at Wave 1	Estimated number of persons who were nonusers at both Wave *w* and Wave 1	Estimated number of persons who were users at Wave *w* although nonusers at Wave 1	
User at Wave 1	Estimated number of persons who were nonusers at Wave *w* although users at Wave 1	Estimated number of persons who were users at both Wave *w* and Wave 1	
Total			

Note that the cells of the contingency table are based on the characteristics of persons who responded at both waves.

- Logistic regression estimating Wave *w* e-cigarette use from both Wave 1 and Wave *w* characteristics. This might include Wave 1 e-cigarette use status (e.g., general user, experimental user, or nonuser), opinion on tobacco use (as identified at Wave 1) of people important to the respondent, education level at Wave *w*, and demographic information. This analysis can be thought of as estimating the probability that a person uses e-cigarettes at Wave *w* from that person's Wave 1 and Wave *w* covariates.

To perform such longitudinal analyses, merge the wave-specific files by PERSONID and then limit the resulting file to records that are common to both files. Use the Wave *w* full-sample single-wave weights and replicate weights if replication methods are to be used for variance estimation purposes.

5.4.3.2 Longitudinal Analyses Using Data From Preceding Waves in Estimating a Later Wave Outcome

Analyses estimating Wave *w* outcomes/behaviors/characteristics from data available from preceding waves are also longitudinal analyses. However, they involve persons who participated in all these waves (as opposed to the analyses discussed in Section 5.4.3.1, which do not use information from the waves between the first wave of a cohort and Wave *w*). For these analyses, use the Wave *w* all-waves weight. Also, use this weight if it is desired to estimate Wave *w* outcomes from characteristics from an earlier wave other than that from the first wave of a cohort (for example, Wave *w-1*). This weight is the appropriate weight to use in the analysis of Wave *c* cohort data including shadow youth recruited at Wave *c* since they will only have data for waves after the first wave of the cohort.

Examples of such analyses are as follows:

- Logistic regression estimating a person's e-cigarette use at Wave *w* from the person's usage status and demographic characteristics at Waves 1 through *w-1*.

- Mediation analysis estimating a person's e-cigarette use at Wave *w* from the person's usage and demographic characteristics at Wave 1, mediated by peer opinions at the intermediate waves (Waves 2 through *w-1*).

- Logistic regression performed on daily cigarette users at Wave 1, relating cigarette usage at Wave *w* to the history of e-cigarette use at Wave 1 through Wave *w-1*.

- Survival analysis estimating time to first trying e-cigarette at Wave *w* for a person who has never tried e-cigarette before, relating the occurrence to peer pressure and social media exposure at Wave 1 through Wave *w*-1.

To perform such longitudinal analyses, merge the wave-specific files by PERSONID and then limit the resulting file to records that are common to all files. Use the Wave *w* full-sample all-waves weights and replicate weights if replication methods are to be used for variance estimation purposes.

5.4.3.3 Cross-sectional Analyses Comparing Different (or Partially Overlapping) Sets of Persons Between Waves

Some researchers may want to use the PATH Study data to answer questions that relate to the cross-sectional populations at the different waves. As explained in Section 5.4.2, there are subtle differences between the PATH Study target population at Wave *w* and the CNP of the United States as of that same point in time. Therefore, any comparisons of these cross-sectional populations using the PATH Study data are approximations.

The PATH Study was not designed to answer questions that relate to the cross-sectional population at different waves. That said, there are situations where cross-sectional estimation may be called for. The following are some examples where cross-sectional comparisons between waves may be desired.

- Comparing the estimated percentage of 7th graders who have tried e-cigarettes at Wave 1 with the estimated percentage of 7th graders who have tried e-cigarettes at a later wave. Most of the 7th graders at Wave 1 are different persons than the 7th graders at a later wave because the targeted interval between waves is 1 year, so a longitudinal analysis is not appropriate (longitudinal analyses are based on data collected from the same persons followed over time).

- Comparing the estimated percentage of 18- to 24-year-olds who use e-cigarettes at Wave 1 with the estimated percentage of 18- to 24-year-olds who use e-cigarettes at a later wave, Wave *w*. There may be some persons who are ages 18 to 23 at Wave 1 and still in the 18-24 age group at Wave *w*. However, generally, the persons who are 24 at Wave 1 will not be in the age group at Wave *w*, and the persons age 18 at Wave *w* will not be in the comparison group for Wave 1. As with the first example, the analysis is not based data collected from the same set of persons followed over time, so it is not longitudinal in nature.

To create cross-sectional estimates for comparing waves, take the following steps:

- Identify the appropriate wave-specific weight variables for the comparison by consulting Table 14. Note that for each type of weight there are separate weight variables for youth and adults, but this distinction is ignored here for simplicity;

- Rename all wave-specific variables, including the weight variables, to obtain a single common name for each set of comparable variables;

- Create a wave indicator variable;
- Concatenate or "stack" data files from each wave to form a single file with one record per respondent per wave in which they provided data.

The subsequent analyses must include the newly created wave indicator variable and the design correctly specified in a software package designed to capture sample variability as described in Appendix A. Even though there may not be complete overlap between the two sets of respondents, there are still correlations between the two groups that should be reflected due to potential partial overlap and because some persons may be in the same PSUs. This correlation serves to reduce the estimated variance of the comparison. Manipulating the variables, records, and files as described above and using the appropriate variance estimation methods will correctly reflect these correlations.

Users are advised not to compare groups at two different waves (or indeed, two subgroups in the same wave) by computing separate confidence intervals for the two estimates and determining if they overlap. Such an analysis would be incorrect even for simple random samples[34] and would result in misleading inferences when applied to a complex design such as used for the PATH Study. This is because such an approach would not account for the correlation between the two groups caused by the complex sample design.

5.4.3.4 Cross-sectional Analyses Describing Findings for a Single Wave Without Comparison to Other Waves

Prior to the release of the Wave 2 PATH Study data, the only analyses possible were cross-sectional analyses of the Wave 1 data. With the release of later waves of data, most of the analyses conducted are anticipated to be longitudinal, developing estimates related to a Wave *w* outcome from characteristics from previous waves (as described in Sections 5.4.3.1 and 5.4.3.2). Cross-sectional analyses comparing different (or partially overlapping) sets of persons between waves may also be undertaken (as described in Section 5.4.3.3). However, there may be some situations in which analysis of later-wave data alone is of interest, for example, to analyze new content introduced in that wave.

[34] See Schenker and Gentleman (2001).

When performing cross-sectional analyses describing findings for a single wave without comparison to other waves, use the data files and weights identified for cross-sectional use in Table 14. For such analyses, the populations represented at Waves 1 and 4 will be the CNP at the time of those waves. The populations represented at Waves 2 and 3 would be, strictly speaking, slightly different than the CNP at those waves, as discussed earlier (e.g., at Wave 2, the estimates represent those in the CNP at Wave 1 who were residents of the U.S. and not incarcerated at Wave 2). However, the estimates would be suitable to serve as reasonable approximation at those waves of the CNP at that time and could be described as such.

6. Data Files

6.1 Data Structure

As discussed in Section 1.2, the adult data file and youth/parent data file for a given wave have similar content. They both include the questionnaire items from the interview and some derived variables such as tobacco-use definitions. The Wave 1 and Wave 4 data files also include census region. .

Starting with Wave 3, a new type of data file was developed in order to simplify the derivation of certain variables in the adult and youth data files at follow-up waves. This new data file is called the Ever/Never Reference data file.

Adult: The adult data file contains one record for every adult interview completed in the given wave.

Youth/Parent: The youth/parent data file contains one record for every youth interview completed in the given wave and also contains responses to the brief parent interview about the youth who completed the youth interview. Not every youth interview has a corresponding parent interview because a parent/guardian can refuse to complete the parent interview but give permission for his/her youth to complete the youth interview. A youth may also have a completed parent interview in one wave but not the next and the same parent may not complete the interview at each wave. A single parent could have completed the parent interview for multiple youths within a household. In this instance, there are certain questions in the parent interview that are only asked once. The data for questions that were not asked in additional interviews with the same parent have been copied from the record of the first interview completed by that parent. For reference, questions that are not repeated in these instances in the parent interview are identified in the parent annotated instrument (combined in the same file as the youth annotated instrument). For example, in Wave 1, these are identified in Boxes R01_PTR01 and R01_PNR01.

Ever/Never Reference: The Ever/Never Reference data file is an auxiliary file that provides respondents' ever/never statuses for tobacco use as of each wave. The objective of this data file is to consolidate information across waves, to ascertain whether respondents ever reported using any tobacco products. Consolidating this information into a single data file helps simplify the algorithms

to derive tobacco use variables in subsequent waves. Without the consolidation of this information, users would need to merge data across multiple waves to determine if respondents reported ever using any tobacco products.

The Wave 3 Ever/Never Reference data file contains one record for each Wave 1 adult, youth, and shadow youth respondent. It includes a set of marker derived variables (MDVs), as well as flag variables for shadow youth and respondents who skipped one or more waves of the study. Each MDV indicates respondents' ever/never status of tobacco product use, as of the most recent wave he or she participated in. The MDV variables in the Wave 3 Ever/Never Reference data file are used to derive ever/never status variables in the Wave 4 adult and youth data files. (The Wave 4 Ever/Never Reference data file will be released when the next wave of PATH Study data are available.)

The intent of the MDVs in the Ever/Never Reference data file is to simplify derivation of tobacco use variables. As such, neither the Ever/Never Reference data file nor the MDVs it contains are recommended for any analyses.

6.2 Record Identifier

The variable named PERSONID serves as the unique identifier for each record in the adult and youth/parent data files. In the adult data files, each PERSONID represents a unique adult respondent. In the youth/parent data files, each PERSONID represents a unique youth respondent. The PERSONID will remain constant throughout the PATH Study even as a youth respondent ages into adulthood. Each PERSONID begins with a "P" and is followed by a randomly assigned nine-digit number that does not include any direct or indirect references to personally identifiable information or geographic location.

6.3 Excluded Variables

Some specific questionnaire items are not provided in the data files. This may be for one of two reasons:

1. Some items are excluded to reduce the risk of disclosing a respondent's identity. All direct identifiers such as names, addresses, and phone numbers are excluded from the files, but other specific questionnaire items are also excluded for this reason.

2. Some items were collected or created in the instrument for operational purposes only. These include items derived throughout the instrument for routing purposes, including items that indicate types and levels of tobacco usage. (See the appendices of the annotated instruments for the full list of tobacco-use variables used for routing purposes.) Variables indicating types and levels of tobacco usage that are better suited for analysis were derived and provided with the data files.

Appendix B contains the full list of questionnaire items excluded from the data files.

6.4 Variable Names

The PATH Study variable naming convention was designed to provide context for each variable based on the content being represented by the item and thus to facilitate the identification and classification of items for analyses. The variable naming convention also includes an indicator that specifies the longitudinal data collection wave.

The naming convention applies to all variables on the data files, with the exception of record identifiers and the variables that reflect the variance structure. It is anticipated that these items will retain the same variable name throughout the course of the PATH Study.

Variables on the PATH Study data files are named using multiple components, each of which is separated by an underscore "_" character for readability. The naming convention is based on a syntax methodology that begins with broader modifiers, moving to more specific modifiers. The basic structure of the naming convention is given in Figure 3. This structure creates variables that are platform independent (no unusual characters) and easy to read, and allows variables to be visually identifiable as related. (The general structure applied across all PATH Study file types aims to prevent the situation where a variable of the same name has different contents in different files. This helps users to avoid unintentionally overwriting data when merging records from two or more files (e.g., when merging adult and youth/parent interview data files). Users may stack (append)

datasets for analysis including both adults and youth, after renaming the variables of interest including the weight and replicate weight variables.)

Figure 3. PATH Study variable name components

The components of the variable names are:

- **Wave Indicator:** Each variable name begins with a wave indicator, represented by three characters. The first character is "R", which is followed by a two-digit zero-filled number indicating the data collection wave, beginning with "01" for Wave 1. As such, all variables in the Wave 1 data files other than PERSONID, VARSTRAT and VARPSU have the prefix "R01".

  All derived and imputed variables include the additional character "R" (indicating that the variable is a recode) after the wave indicator. As such, all derived and imputed variables on the Wave 1 data files have the prefix "R01R". Similarly, all variables created from an external source include the additional character "X" so that all such variables on the Wave 1 data files have the prefix "R01X".

  All marker derived variables in the Ever/Never Reference Files include the additional character "M" after the wave indicator. As such, all marker derived variables in the Wave 3 Ever/Never Reference Files have the prefix "R03M," and so on for each wave.

- **Modifiers**

  - **PATH ID:** A PATH ID represents a unique question in an instrument. All interview variables are named using the PATH IDs that were assigned to questions in the instrument specifications. PATH IDs are represented in the format: @$####, where @ represents the instrument type (A for Adult, P for Parent, and Y for Youth), $ represents the item content (e.g., C for cigarettes), and #### is a numeric key that is unique within each instrument. The numbers in the PATH IDs do not correspond to any specific ordering of questions; they were simply assigned during the instrument design process based on availability. Some variable names include suffixes to the PATH ID to differentiate between tobacco product types when the same question was repeated for different tobacco products. For example, variable R01_AG1005TC corresponds to the question with PATH ID AG1005 when it was asked about traditional cigars in the Wave 1 adult interview. In some instances, there are questions with similar content that are asked in different sections or in different instruments. In such cases, the number identifier in the PATH ID is the same across both sections/instruments whereas the instrument/content indicator is different. For example, in Wave 1, the question with PATH ID AC1005 (lifetime use of cigarettes) in the cigarette

section is also asked about e-cigarettes in the question with PATH ID AE1005 (lifetime use of e-cigarettes).

– **Interview Identifier:** Where appropriate, administrative, management, processing, derived, and imputed variables include an indicator of the interview to which the information pertains: "A" (adult interview), or "Y" (youth/parent interview). For example, the full-sample weight for analyzing data from the Wave 1 adult interview is in the variable named R01_A_PWGT; the corresponding weight for analyzing data from the Wave 1 youth or parent interview is in the variable named R01_Y_PWGT.

– **Mnemonic:** All administrative, management, processing, derived, external, and imputed variables are named using mnemonics. Mnemonics can be full words, phrases, or abbreviations to represent the content of the variable, and can range from the very simple to the complex. For example, the variable indicating the youth's body mass index at Wave 1 is R01R_Y_BMI. In other cases, the mnemonic is actually the PATH ID. For example, the adult variable indicating the age range when first alcoholic drink was consumed is recoded from the questions AX0074 (age in years) and AX0270 (age range, if age in years not given) because all age in year variables were excluded from the PUFs; the recoded variable is named R01R_A_AX0074. Variables created from combining parent and emancipated youth responses to the same question are given a hybrid mnemonic. For example, the variable indicating the youth's overall health status is recoded from PT0035 (parent response) and YT0035 (emancipated youth response); the recoded variable is named R01R_Y_PY0035.

  The tobacco-use variables have components in the mnemonic that indicate the status, intensity, timeframe, and product. For example, the variable indicating whether an adult was a former established past 12 months cigarette user at Wave 1 is R01R_A_FMR_ESTD_P12M_CIGS. Table 15 includes a list of the tobacco-use mnemonic components.

Table 15. Tobacco-use variable name mnemonic components[35]

Type/Attribute	Abbreviation	Description
Status	CUR	Current
	EDY	Every Day
	SDY	Some Days
	EDSD	Every Day or Some Days
	NVR	Never
	FMR	Former
	NEW	New
	EVR	Ever

[35] The maximum length for a variable name in the PATH Study data files is 32 characters. Current and recent versions of most statistical software packages are able to accommodate variable names with 32 (or more) characters.

Table 15. Tobacco-use variable name mnemonic components (continued)

Type/Attribute	Abbreviation	Description
Intensity	ESTD	Established User
	EXPR	Experimental User
	THRSHLD,[36] THRSH	Lifetime Threshold of Use
Timeframe	12MA	12 Months Ago (one year ago)
	P12M	In Past 12 Months
	P30D	In Past 30 Days
Product	CIGS	Cigarettes
	CIGSMFG	Cigarettes, Manufactured
	CIGSRYO	Cigarettes, Roll Your Own
	DISSBL	Dissolvables
	ECIG	E-Cigarettes
	EPRODS	E-Products
	ECIGAR	E-Cigars
	EPIPE	E-Pipe
	EHOOK	E-Hookah
	GFILTR	Filtered Cigars
	GRILLO	Cigarillos
	GTRAD	Traditional Cigars
	CIGAR	All Cigars
	HOOK	Hookah
	NRT	Nicotine Replacement Therapy
	PIPE	Pipe
	RX	Prescription Therapy
	SMKLS	Smokeless
	SNUS	Snus Pouches
	BIDI	Bidis
	KRETEK	Kreteks
	TOB	All Tobacco Products

- **Source:** As additional sources of external data are added, a modifier for the source will be included. For example, consider item R01X_CB_REGION, which represents census region in which the respondent was sampled; CB identifies the information source as the Census Bureau.

- **Sub-item Identifier:** Some interview questions have multiple components for a single PATH ID, such as choose-all-that-apply items or dates where responses are recorded in multiple variables. In such cases, the variable name includes a sub-item identifier to differentiate components. For example, consider item R01_AG9009, about cigarillos and filtered cigars. The question and response options are shown in Figure 4 for reference.

[36] The abbreviation THRSHLD was only used in tobacco-use variable names in the Wave 3 adult data file. The abbreviation THRSH is used in tobacco-use variable names in all other data files.

Figure 4. Example of question and response options

Which of the following kinds of cigarillos or filtered cigars have you smoked[as blunts]? Choose all that apply. The kind…		
R01_AG9009_01	1	With a plastic or wooden tip
R01_AG9009_02	2	With a filter (like a cigarette filter)
R01_AG9009_03	3	Without a tip or filter
	-8	DON'T KNOW
	-7	REFUSED

The respondent could choose multiple responses for this item. Because there are three sub-items for R01_AG9009, there are three variables (R01_AG9009_01 through R01_AG9009_03), each of which corresponds to one of these sub-items. The variables for the sub-items correspond to the order of the responses presented in the question. Therefore, if a respondent chose "With a plastic or wooden tip" and "Without a tip or filter" as answers to R01_AG9009, only the variables R01_AG9009_01 and R01_AG9009_03 contain the value, indicating the respondent chose that response option (typically this value is 1).

6.5 Variable Labels

All variables on the data files include a label that briefly explains the content of the item. For clarity, consistency, and usability, labels are assigned to each variable using a standard convention. The components of the variable labeling convention are shown in Figure 5.

Figure 5. PATH Study variable labeling convention

Variable Name	:	[DERIVED | IMPUTED | IMPUTATION FLAG | MARKER -]Brief Description

For all interview, administrative, and management variables, the labels begin with the variable name followed by a colon and a brief description of the content. For derived, imputed, or imputation flag variables, the labels begin with the variable name followed by a colon, the respective phrase "DERIVED", "IMPUTED", or "IMPUTATION FLAG", a spaced hyphen, and then a brief description of the content. For marker derived variables in the Ever/Never Reference data files, the labels begin with the variable name followed by a colon, the key word "MARKER," a spaced hyphen, and then a brief description of the content.

6.6 Value Labels

The labels associated with variable values are provided. All value labels include both the data value as well as the description of the value. For example, a variable for an item with yes/no responses has "1 = Yes" and "2 = No" as value labels.

Interview Variables

The value labels for all interview variables match the text of the response option as seen by the respondent.

Recoded Values

Certain variables have new recoded values that were not presented as response options during the interview. Recoded values are assigned to items with Other-Specify response options. The label for any recoded value is prefixed with the term "RECODE:" for explicit identification.

Values for Derived, Imputed, and External Variables

The labels for all created values for derived, imputed, and external variables include the coded value and a description of the value. If a value could not be assigned, it is coded as a missing value. Please see Section 6.7 for a description of the missing value codes.

6.7 Missing Values

The data files include missing values, as currently defined in Table 16. The user is advised to refer to the descriptive frequencies in the documentation accompanying the data files to ascertain the levels of missing data in variables of interest.

Table 16. PATH Study missing value codes and their descriptions

Value	Definition*	Applicable data types	Applicable variable types	Description*
-1	Inapplicable	Numeric, Character	Interview	This value is coded to identify items that were validly skipped on a route specific to a respondent.
-5	Improbable response removed	Numeric	Interview	This value is coded to represent an improbable numeric answer provided by a respondent.
-6	Unassigned PERSONID	Character	Interview	This value is coded when a new person is added as a spouse or other guardian in parent interviews for Wave 2 and beyond.
-7	Refused	Numeric, Character	Interview	This value is coded when a respondent answered a question as "Refused."
-8	Don't know	Numeric, Character	Interview	This value is coded when a respondent answered a question as "Don't know."
-9	Missing – Not ascertained	Numeric, Character	Interview	This value is coded when data were supposed to be collected on a route specific to a respondent but the collection did not occur due to a skip error in the instrument.
-99933	Missing due to respondent not being in previous wave	Numeric, Character	Derived	This value is coded when a value in a derived variable is missing because a respondent was a shadow youth in a previous wave. It is also coded when a value in a derived variable is missing for respondents added in a replenishment wave.
-99999	Missing due to data not ascertained on one or more component variables	Numeric, Character	Derived	This value is coded when a value in a derived variable is missing because data were "Not ascertained" (as described above) in one or more component variables.
-99988	Missing due to a don't know response on one or more component variables	Numeric, Character	Derived	This value is coded when a value in a derived variable is missing due to a "Don't know" response in one or more component variables and none of those variables were "Not ascertained" (as described above).
-99977	Missing due to a refused response on one or more component variables	Numeric, Character	Derived	This value is coded when a value in a derived variable is missing due to a "Refused" response in one or more component variables and none of those variables were missing due to a "Don't know" response or because they were "Not ascertained" (as described above).
-99955	Missing due to an improbable response on one or more component variables	Numeric	Derived	This value is coded when a value in a derived variable is missing due to an improbable response in one or more component variables and none of those variables were missing due to "Don't know" or "Refused" responses or because they were "Not ascertained" (as described above).

Table 16. PATH Study missing value codes and their descriptions (continued)

Value	Definition*	Applicable data types	Applicable variable types	Description*
-99966	Missing due to missing data on a prior wave for one or more component variables	Numeric, Character	Derived	This value is coded when a value in a derived variable is missing because one or more component variables from any previous wave is coded as a missing value. It is also coded when a respondent did not participate in one or more previous waves.
-99911	Missing due to an instrument skip pattern for one or more component variables	Numeric, Character	Derived	This value is coded when a value in a derived variable is missing due to items that are not applicable/validly skipped for a respondent in one or more component variables and none of those variables were missing due to the other reasons listed above.
-97777	Missing due to data removed per respondent request	Numeric, Character	Interview, Derived	This value is coded when a respondent requested data removal.

* A component variable is any variable used in the creation of a derived variable. Unless otherwise stated, "component variables" are from the current wave.

An example of the implementation of these missing values codes can be seen in the "ever" status variables. An "ever" status identifies whether a respondent has ever used a tobacco product (e.g., derived variables named R0*w*R_*x*_EVR_*aaaaa*, where *w* indicates the wave, *x* indicates "A" or "Y" respectively, and *aaaaa* is an abbreviation for a tobacco product). When respondents did not participate in a prior wave but indicated a "never" status at Wave 1 (baseline wave) and a "not in the past 12 months" status for all other waves in which they participated, an "ever" status was not conclusively identified. In such instances, the derived "ever" status variables have been coded as missing, with a value -99966.

6.8 Outlier Values

The PATH Study adult and youth instruments were self-administered, and although there were checks throughout the instrument to query potentially illogical or unlikely responses, the number of such checks was limited in order to minimize frustration on the part of the respondent. Moreover, these checks were designed to warn the respondent if a response was unlikely, but the respondent could continue through the instrument without changing the response. In order to maintain the confidentiality afforded by the ACASI instrument, it was important to minimize the number of

instances where the respondent might need to ask the field interviewer for assistance in overcoming a check that required correction of data before continuing. As a result, the PATH Study questionnaire data include outliers, i.e., values that lie outside the expected range of data. There were two types of outliers, which were handled in two different ways.

First, there are outliers that are logically impossible. For example, a few respondents reported doing something at an age that was older than their current age (such as a respondent who was age 37 reporting initiation of cigarillo use at age 39). For these clear instances where a response was logically impossible, the outlier was removed from the dataset, replaced with a value of -5, and labeled as "Improbable response removed." Please see Appendix C for a full list of interview variables with coded outlier values.

Second, there are outliers that are unlikely but logically possible. For example, a few respondents reported sending more than 1,000 text messages a day, which is unlikely but not impossible. In the absence of established upper or lower limits for such unlikely but logically possible values, they were retained in the dataset. Therefore, analysts are advised to review carefully the distributions of the variables used in all analyses, and to consider the impact of these values on any findings.

6.9 Age Collection, Calculation, and Inconsistencies Across Waves

The sections below describe how age information is collected in the PATH Study (Section 6.9.1), how the age variables are calculated (Section 6.9.2), and how inconsistencies in ages across waves can result (Section 6.9.3). Analytic considerations in light of such age inconsistencies are provided in Section 6.9.4.

6.9.1 Collection of Age Information

All adults sampled in recruitment waves were asked to provide their date of birth as part of the consent process. If the adult refused or reported that they did not know their date of birth, they were asked for an exact age (at the time of the consent). If the adult also refused or reported that they did not know their exact age, they were asked to choose a response from among specified age ranges. If the adult refused to provide or did not know any of this information, the consent process was ended and the adult was not included in the PATH Study.

Parents of youth sampled in recruitment waves were asked to provide the date of birth for their child. If a parent refused or reported that they did not know the sampled youth's date of birth, the youth was not included in the PATH Study. The reason for the more stringent requirement for youth respondents was to ensure that the appropriate consent and assent procedures were followed for study participants under age 18.

Because the PATH Study includes youth under age 18 and asks about illegal and/or sensitive behaviors, ensuring that the study correctly identifies the age of each respondent prior to a follow-up interview is of high priority. This requires confirming or asking again about age at follow-up waves for all respondents, as described below.

The process for confirming or collecting age at follow-up waves for adults depended on respondent type:

- If continuing adult respondents provided their date of birth at a previous wave, they were asked to verify the date of birth.
- If continuing adult respondents did not provide their date of birth at a previous wave, they were asked for their date of birth.
- Aged-up adults were asked to verify the date of birth previously provided by their parent.

In any of the above scenarios, if an adult respondent indicated that their date of birth provided at a previous wave was incorrect, they were asked for the correct date of birth, with follow-up questions similar to those used at the recruitment wave if the respondent refused or reported that they did not know the information. If an adult respondent did not provide updated age information, the interview for the respective wave proceeded.

For youth respondents, the age information was confirmed in follow-up waves by asking the parent to verify the date of birth that was provided for their child at a previous wave. If the parent reported that the information was incorrect, they were asked to provide the correct date of birth. If a parent did not provide updated information, the youth interview was coded as nonresponse for that wave.

6.9.2 Calculating Interview Ages

For each wave, the calculated interview age is in the variable R0*w*R_*x*_AGE, where *w* indicates the wave number and *x* indicates "A" for adults or "Y" for youth.

For all respondents who provided a date of birth at recruitment, their age as of the date of the respective interview was calculated by subtracting the date of birth from the interview date for the recruitment wave. If an exact age was provided, that value was used as the age for the recruitment wave. If only an age range was provided, the calculated age for the recruitment wave is missing. The calculated age was then used to create an age range variable.

For follow-up waves, age at each wave was calculated as follows:

- For respondents who verified the date of birth provided at a previous wave, their age as of the date of the respective follow-up interview was calculated by subtracting the confirmed date of birth from the date of the follow-up interview.
- If a respondent revised their date of birth at a follow-up wave, their age as of the date of the respective follow-up interview was calculated using the revised date of birth.
- If a respondent provided an exact age at a follow-up wave, that value was used.
- If a respondent only provided an age range at a follow-up wave, the exact interview age is missing. However, respondents were routed appropriately through the instrument at each respective wave for the given age group.

The age range corresponding to this calculated age is in the variable R0*w*R_A_AGECAT7 for Wave 1 to Wave 3, and R04R_A_AGECAT6 for Wave 4 adults and R0*w*R_Y_AGECAT2 for youth.

6.9.3 Potential for Age Inconsistencies Across Waves

Logically, a respondent with age z at Wave *w* is expected to have an age of z, z+1, or z+2 at Wave *w+1*, depending on the date of the Wave *w* interview and the follow-up interview in relation to the respondent's date of birth. Although every effort is made to complete follow-up interviews at approximately 1-year intervals, participants have the opportunity to respond until the end of each wave. Therefore, a respondent could have two birthdays between interviews, one occurring soon

after their Wave *w* interview and another soon before their Wave *w+1* interview. As such, a difference of 2 years between Wave *w* and Wave *w+1* interview ages is considered acceptable.[37]

However, if in a follow-up wave a respondent provided a revised date of birth or age that constitutes a logical inconsistency from previously recorded age(s), the longitudinal age progression at any follow-up wave may fall outside of the expected range.

A change to information that results in inconsistent interview ages between waves is problematic for analysis of longitudinal data. Because respondents are expected to increase in age by no more than 2 years between interviews depending on the timing of their birth dates and their interview dates, the problematic cases are those where a respondent revised their information such that at Wave *w+1* they were younger or more than two years older than they appeared to be at Wave *w*.

So that respondents with inconsistent ages across waves are readily apparent, the dichotomous variable R0*w*R_*x*_AGE_CHECK is included on the adult and youth data files after Wave 1 (where *w* indicates any wave after Wave 1 and *x* indicates "A" or "Y" as appropriate) with the following categories:

1 = Age is within expected range

2 = Age is not within expected range.

Some respondents with inconsistent ages across waves will not appear to have inconsistent age ranges; because the ages are within the same age range or result in a change from one range to the next higher range. The age-check variable is included on the files, even though single-year of age is not provided, to identify all suspect cases that analysts may want to exclude or categorize separately as discussed in the next section.

[37] Some respondents participate in the PATH Study in non-consecutive waves. In such instances, a 2-year difference for each wave is included in the total acceptable interval between the waves completed by respondents. For example, if a respondent participates in Wave 1, does not participate in Wave 2 and participates in Wave 3, a difference of 4 years (between Wave 3 and Wave 1) is considered acceptable.

6.9.4 Analytic Considerations

Although the number of cases with inconsistent ages across waves is small, users may want to consider the influence these cases have on their analyses:

- For analyses that predict an outcome at a follow-up wave based on data from an earlier wave, the age at the earlier wave is probably the more appropriate choice (for example, when predicting first use of e-cigarettes among Wave 1 nonusers of e-cigarettes, the Wave 1 age would be employed).
- For analyses that include only a single wave of data (for example, an analysis of Wave 2 e-hookah use), the age at the wave being analyzed is probably the more appropriate choice.

For these and other types of analysis, it is suggested that a sensitivity analysis be conducted to determine if including respondents who have an age that is not within the expected range makes a difference in the results.

6.10 Considerations for Pooling Data

Users interested in pooling PATH Study data with data from other sources should first consider the advantages and disadvantages of any such analysis. Pooling data may increase statistical power, particularly for subgroups; however, independent studies can differ in their measurement of key constructs, timing of the study, and target population. Measures of certain constructs may even differ across assessment periods within the same study. Analysts should use weights and/or design variables appropriately to ensure that the complex design features of the PATH Study (and potentially other data sources) are reflected in estimates based on pooled data.

7. Linking Files

The PATH Study public use files include a master linkage data file that indicates in which file(s) a participant has data. The linkage file can help analysts identify which files contain data for a particular participant (or set of participants) and will be extended with each wave. This file is described in more detail in Section 7.1.

Data files may also be linked together across waves. This type of linkage is described in Section 7.2. Sections 7.3 and 7.4 provide some general file merging notes and recommendations, and information about the case identification variable created by ICPSR.

7.1 Master Linkage Data File Structure

The master linkage data file includes a unique record for each PERSONID corresponding to each PATH Study participant:

- Wave 1 adult and youth interview respondents.
- Wave 1 shadow youth, even though no data exist for shadow youth in the PATH Study Wave 1 data files.
- Wave 4 adult and youth interview respondents from the replenishment sample.
- Wave 4 shadow youth from the replenishment sample, even though no data exist for shadow youth in the PATH Study Wave 4 data files.

The records for shadow youth are included in the master linkage data file to account for participants who aged up to youth at a follow-up wave.

The master linkage data file will be maintained as a standalone dataset that will be updated with each data file release. The number of records in the master linkage data file is defined by the adult, youth, and shadow samples established at Wave 1 and future replenishment efforts. The number of variables in the master linkage data file will increase as each new file with respondent-level data is released. The number of variables in the master linkage data file will also increase as data for future waves become available.

The master linkage data file includes indicator variables for the availability of interview data, weights, and biomarker data. Interview indicators are named WAVE*w*_INTERVIEW where *w* denotes the wave (e.g., 1, 2, 3, 4). Biomarker indicators are named WAVE*w*_*descriptor* where *w* denotes the wave and *descriptor* denotes the data file contents (e.g., URINE_COLLECTED, BLOOD_COLLECTED, ASSPEC_LAB, HSCRP_LAB).

To identify the wave in which each PATH Study participant was sampled, the master linkage data file includes a variable named RECRUITMENT_WAVE. The master linkage data file also includes variables to identify each cohort, as described in Chapter 2. These variables are named WAVE*c*_COHORT, where *c* indicates the wave at which the cohort was established. Please refer to the master linkage data file codebook for a description of the variables and their values.

Table 17 below illustrates possible response scenarios indicated by the comparison of the variable WAVE1_INTERVIEW to variables added for subsequent waves for the PATH Study sample established at Wave 1.

Table 17. Possible response scenarios when the variable WAVE1_INTERVIEW is compared to variables added for subsequent waves

Description/Scenario	WAVE1_INTERVIEW	WAVE*w*_INTERVIEW*
Adult across waves	1 = Adult	1 = Adult
Adult in Wave 1, no response in Wave *w*	1 = Adult	9 = Nonresponse
Youth in Wave 1, Adult in Wave *w*	2 = Youth	1 = Adult
Youth across waves	2 = Youth	2 = Youth
Youth in Wave 1, no response in Wave *w*	2 = Youth	9 = Nonresponse
Shadow youth in Wave 1, Adult in Wave *w*	3 = Shadow Youth (No data)	1 = Adult
Shadow youth in Wave 1, Youth in Wave *w*	3 = Shadow Youth (No data)	2 = Youth
Shadow youth across waves	3 = Shadow Youth (No data)	3 = Shadow Youth (No data)
Shadow youth in Wave 1, no response in Wave *w*	3 = Shadow Youth (No data)	9 = Nonresponse

*Note that *w* indicates any wave after Wave 1.

PATH Study participants sampled at Wave 4 do not have records in any data files released for Waves 1 through 3. As such, for records that represent participants sampled at Wave 4 in the master linkage data file (RECRUITMENT_WAVE = 4), the value "0 = Not in sample" is assigned to all variables that correspond to data files for Waves 1 through 3.

Table 18 below illustrates example response scenarios indicated by the comparison of the variables RECRUITMENT_WAVE, WAVE1_COHORT, WAVE4_COHORT, and WAVE4_INTERVIEW for the full PATH Study sample at Wave 4, including the replenishment.

Table 18. Example response scenarios when the variables RECRUITMENT_WAVE, WAVE1_COHORT, WAVE4_COHORT, and WAVE4_INTERVIEW are compared

Description/ Scenario*	RECRUITMENT_WAVE	WAVE1_COHORT**	WAVE4_COHORT**	WAVE4_INTERVIEW
Adult sampled in Wave 1, responding in Wave 4 and in the Wave 4 Cohort	1	1	1	1 = Adult
Adult sampled in Wave 4, responding in Wave 4	4	2	1	1 = Adult
Adult sampled in Wave 1, responding in Wave 4 and not in the Wave 4 Cohort	1	1	2	1 = Adult
Adult sampled in Wave 1, no response in Wave 4	1	1	2	9 = Nonresponse
Youth sampled in Wave 1, responding Youth in Wave 4 in the Wave 4 Cohort	1	1	1	2 = Youth
Youth sampled in Wave 4, responding in Wave 4	4	2	1	2 = Youth
Youth sampled in Wave 1, no response in Wave 4	1	1	2	9 = Nonresponse
Shadow youth sampled in Wave 1, responding Youth in Wave 4 and in the Wave 4 Cohort	1	1	1	2 = Youth
Shadow youth sampled in Wave 4, responding in Wave 4	4	2	1	3 = Shadow Youth (No data)

*Scenarios in this table are examples of response patterns in the data and are not meant to be comprehensive.

** 1=Yes, 2=No.

7.2 Linking Data Files Across Waves

The unique identifier in all files is PERSONID. This allows users to easily merge the data from multiple waves for longitudinal analyses.

It is not necessary to use the master linkage data file to merge data across waves. For example, the Wave 1 adult data file can be merged directly with the Wave 2 adult data file using PERSONID. The primary utility of the linkage data file is to help analysts identify the different types of data available for a particular respondent or set of respondents as identified by PERSONID, both within and across data collection waves. For example, the linkage file can be used to identify the subset of respondents who completed a youth interview in Wave 1 and an adult interview in Wave 2. The resulting subset of PERSONIDs identified using the linkage data file can be used to merge records from multiple files.

Data for PATH Study participants sampled in Wave 4 cannot be linked to other waves at this time. Longitudinal analyses can be performed for these participants when data from follow-up waves beyond Wave 4 are available.

7.3 File Merging Notes and Recommendations

Data users are advised to avoid merges that retain all variables from multiple files because the adult and youth/parent data files each have a large number of variables. The process of merging data files with large numbers of variables may not be resource intensive, but the large size of the resulting (merged) file could significantly increase the time required to run analyses. As such, when merging data files, users are advised to keep only the variables needed from each file to run the intended analyses. This practice keeps the merged file(s) as small as possible, which helps analytic processes run faster.

The records included in the file(s) that result from merging two or more files can vary depending on how the merge is specified. As such, users should carefully consider the set of cases needed for any specific analyses and define the merge accordingly.

7.4 Case Identification Number

ICPSR creates a variable for each of its datasets named CASEID. This variable numbers the cases sequentially so that the file can easily be returned to its original order at any time. This variable should not be confused with any other identification variable previously mentioned. It should not be used to merge files or link records together.

References

Battaglia, M.P., Hoaglin, D.C., and Frankel, M.R. (2013). Practical considerations in raking survey data. *Survey Practice*, 2(5).

Chowdhury, S.R. (2013). Effects of poststratification and raking on precision of MEPS estimates. Proceedings of the Federal Committee on Statistical Methodology Conference.

Chowdhury, S., Khare, M., and Wolter, K. (2007). Weight trimming in the National Immunization Survey. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 2651-2658.

Golder, P.A. and Yeomans, K.A. (1973). The use of cluster analysis for stratification. *Applied Statistics*, 22, 213-219.

Graubard, B.I. and Korn, E.L. (1996). Survey inference for subpopulations. *American Journal of Epidemiology*, 144, 102-106.

Judkins, D.R. (1990). Fay's method for variance estimation. *Journal of Official Statistics*, 6, 223-239.

Judkins, D.R. and Singh, R.P. (1981). Using clustering algorithms to stratify primary sampling units. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 279-284.

Korn, E. and Graubard, B. (2011). *Analysis of Health Surveys*. Hobart, NJ: Wiley.

Krewski, D. and Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics*, 9, 1010-1019.

Lewis, T. (2013). Considerations and techniques for analyzing domains of complex survey data. SAS Global Forum Proceedings, http://support.sas.com/resources/papers/proceedings13/449-2013.pdf.

Lohr, S. (2010). *Sampling: Design and Analysis, 2nd ed.* Boston: Brooks/Cole.

McCarthy, P.J. (1969). Pseudo-replication: Half samples. *Review of the International Statistical Institute*, 37, 239-264.

Rao, J.N.K. and Shao, J. (1993). On balanced half sample variance estimation in stratified sampling. *Journal of the American Statistical Association*, 91, 343-348.

Rao, J.N.K. and Shao, J. (1999). Modified balanced repeated replication for complex survey data. *Biometrika*, 86, 403-415.

SAS Institute Inc. (2018). *SAS/STAT® 15.1 User's Guide*. Cary, NC: SAS Institute, Inc.

Schenker, N. and Gentleman, J.F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, 55(3), 182-186.

Valliant, R. (1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association*, 88, 89-96.

Valliant, R. (2004). The effect of multiple weighting steps on variance estimation. *Journal of Official Statistics*, 20, 1-18.

Wolter, K.M. (2007). *Introduction to Variance Estimation*, *2nd ed.* New York, NY: Springer.

Appendix A
Example Program Code for Generating Popular Statistics

This appendix contains example SAS, SUDAAN, Stata, R, and SPSS program code for generating statistics using the PATH Study data. A couple of notes about the examples:

- Text in italics represents placeholders for actual dataset and variable names.
- These examples use the adult full-sample and replicate weights for Wave 1, but analyses of youth/parent data for Wave 1 or data from other waves would be the same with the respective full-sample and replicate weights.
- In the examples using subdomain analysis, the domain of interest (e.g., Asian youth or adults ages 18-24) is identified by setting *domainvar* = 1 for those cases, and setting *domainvar* = 0 for all other cases.

These examples are provided primarily to illustrate the correct specifications for creating appropriate variance estimates. They are not meant to be exhaustive or to provide instruction for users unfamiliar with a particular software package.

SAS

The following code creates tables including the unweighted frequencies of categorical variables *var1*, *var2*, *var3*, and *var4*, and weighted estimates of population totals and population proportions for each level of those variables (using the weight R01_A_PWGT) along with the standard errors of these estimates (using the replicate weights R01_A_PWGT1 - R01_A_PWGT100) and modified Wilson confidence intervals:

```
proc surveyfreq data=analysis_dataset
varmethod=BRR (fay=0.3);
    tables var1 var2 var3 var4 / cl(type=Wilson);
    weight R01_A_PWGT;
    repweights R01_A_PWGT1 - R01_A_PWGT100;
    title "PROC SURVEYFREQ using BRR-Fay Replication";
run;
```

The following code creates the weighted mean of continuous variable *var5* (using the weight R01_A_PWGT) along with the standard error of that estimate (using the replicate weights R01_A_PWGT1 - R01_A_PWGT100):

```
proc surveymeans data=analysis_dataset
varmethod=BRR (fay=0.3);
    var var5;
    weight R01_A_PWGT;
    repweights R01_A_PWGT1 - R01_A_PWGT100;
    title "PROC SURVEYMEANS using BRR-Fay Replication";
run;
```

The following code fits a linear regression model using continuous variable *respvar* as the outcome variable and continuous variables *cov1* and *cov2* as the predictors. All parameter estimates are weighted (using the weight R01_A_PWGT) and the standard errors are calculated using the replicate weights (R01_A_PWGT1 - R01_A_PWGT100).

```
proc surveyreg data=analysis_dataset
varmethod=BRR (fay=0.3);
    model respvar = cov1 cov2 / solution;
    weight R01_A_PWGT;
    repweights R01_A_PWGT1 - R01_A_PWGT100;
    title "PROC SURVEYREG using BRR-Fay Replication";
run;
```

The following code fits a logistic regression model using dichotomous variable *respvar2* (with values of 0 and 1) as the outcome variable and continuous variables *cov1* and *cov2* as the predictors. All parameter estimates are weighted (using the weight R01_A_PWGT) and the standard errors are calculated using the replicate weights (R01_A_PWGT1 - R01_A_PWGT100).

```
proc surveylogistic data=analysis_dataset
varmethod=BRR (fay=0.3);
    model respvar2 (event='1')= cov1 cov2;
    weight R01_A_PWGT;
    repweights R01_A_PWGT1 - R01_A_PWGT100;
    title "PROC SURVEYLOGISTIC using BRR-Fay Replication";
run;
```

Note that the current SAS/STAT software documentation states that for a complex survey design, analyses for a subpopulation should specify the subpopulation through a `DOMAIN` statement in `proc surveymeans`, `proc surveyreg`, or `proc surveylogistic`, or as the first

dimension in the TABLES statement in `proc surveyfreq` (see SAS Institute Inc. 2018, pp.9819-9820). However, a simplification is possible for analysis of PATH Study data using the replicate weights (Replication is the recommended method of variance estimation for the PATH Study.) **Specifically, the `DOMAIN` statement is not required for analyses using the replication method for variance estimation with the replicate weights provided with the data files.** The appropriateness of this simpler approach has been confirmed with software developers at SAS Institute, Inc.

In some situations, specifying the `DOMAIN` statement when using the replication method can result in inaccurate variance estimates. Thus, when using the replicate weights provided with the PATH Study data, the data should be subset to the subpopulation of interest outside the SAS/STAT software procedures, for example in a separate data step, or within the procedures using a `BY` or `WHERE` statement to produce the correct standard errors for confidence intervals and hypothesis tests even though, in some older versions of SAS, a warning may appear in the program log stating that the method "does not provide a statistically valid subpopulation or domain analysis." For this reason, the example SAS code provided in this User Guide does not include a `DOMAIN` statement.

The following code provides examples of domain analyses in SAS. SAS will produce output for all levels of the domain variable. The Output Delivery System (ODS) statements may also be used to output the domain cases of interest for further processing.

```
proc surveyfreq data=analysis_dataset
varmethod=BRR (fay=0.3);
    table var1 / cl(type=Wilson);
    by domainvar;
    weight R01_A_PWGT;
    repweights R01_A_PWGT1 - R01_A_PWGT100;
    title "Domain Analysis for PROC SURVEYFREQ using BRR-Fay
    Replication";
run;
```

```
proc surveymeans data=analysis_dataset
varmethod=BRR (fay=0.3);
    var var5;
    by domainvar;
    weight R01_A_PWGT;
    repweights R01_A_PWGT1 - R01_A_PWGT100;
    title "Domain Analysis for PROC SURVEYMEANS using BRR-Fay
    Replication";
run;

proc surveyreg data=analysis_dataset
varmethod=BRR (fay=0.3);
    model respvar = cov1 cov2 / solution;
    by domainvar;
    weight R01_A_PWGT;
    repweights R01_A_PWGT1 - R01_A_PWGT100;
    title "Domain Analysis for PROC SURVEYREG using BRR-Fay
    Replication";
run;

proc surveylogistic data=analysis_dataset
varmethod=BRR (fay=0.3);
    model respvar2 (event='1')= cov1 cov2;
    by domainvar;
    weight R01_A_PWGT;
    repweights R01_A_PWGT1 - R01_A_PWGT100;
    title "Domain Analysis for PROC SURVEYLOGISTIC using BRR-Fay
    Replication";
run;
```

SUDAAN

When using Fay's method of BRR in SUDAAN, the BRR-Fay factor must be calculated outside of the software as shown in equation A-1:

$$ADJFAY = 1/[(1-\mathrm{k})^2] \quad \text{(A-1)}$$

When $k = 0.3$, as with the PATH Study, $ADJFAY = 2.040816$.

The following code creates tables including the unweighted frequencies of categorical variables *var1*, *var2*, *var3*, and *var4*, and weighted estimates of population totals and population proportions for each level of those variables (using the weight R01_A_PWGT) along with the

standard errors of these estimates (using the replicate weights R01_A_PWGT1 - R01_A_PWGT100) :

```
proc crosstab data=analysis_dataset filetype=sas design=brr;
     weight R01_A_PWGT;
     repwgt R01_A_PWGT1 - R01_A_PWGT100 / adjfay=2.040816;
     tables var1 var2 var3 var4;
     class var1 var2 var3 var4;
     print / style=nchs tablecell=all;
     title 'SUDAAN proc crosstab using BRR-Fay Replication';
run;
```

By default, SUDAAN generates the logit transformed confidence intervals with the above code. To obtain the Clopper-Pearson small percentage confidence intervals, the SMCONF option is specified in the PROC statement. For example, SMCONF = 25 produces the Clopper-Pearson confidence intervals for percentages less than or equal to 25% or percentages greater than or equal to 75%. For percentages between 25% and 75%, the logit transformed confidence intervals are calculated. To print the Clopper-Pearson confidence intervals for small percentages, the ROWSPCI option is specified in the PRINT statement. The following code provides an example for creating Clopper-Pearson confidence intervals:

```
proc crosstab data=analysis_dataset filetype=sas design=brr
smconf=25;
     weight R01_A_PWGT;
     repwgt R01_A_PWGT1 - R01_A_PWGT100 / adjfay=2.040816;
     tables var1 var2 var3 var4;
     class var1 var2 var3 var4;
     print rowspci/ style=nchs tablecell=all;
     title 'SUDAAN proc crosstab using BRR-Fay Replication and
producing Clopper-Pearson confidence intervals';
run;
```

The following code creates the weighted mean of continuous variable *var5* (using the weight R01_A_PWGT) along with the standard error of that estimate (using the replicate weights R01_A_PWGT1 - R01_A_PWGT100):

```
proc descript data=analysis_dataset filetype=sas design=brr;
     weight R01_A_PWGT;
     repwgt R01_A_PWGT1- R01_A_PWGT100 / adjfay=2.040816;
     var var5;
     print / style=nchs;
```

```
     title 'SUDAAN proc descript using BRR-Fay Replication';
run;
```

The following code fits a linear regression model using continuous variable *respvar* as the outcome variable and continuous variables *cov1* and *cov2* as the predictors. All parameter estimates are weighted (using the weight R01_A_PWGT) and the standard errors are calculated using the replicate weights (R01_A_PWGT1 - R01_A_PWGT100).

```
proc regress data=analysis_dataset filetype=sas design=brr;
     weight R01_A_PWGT;
     repwgt R01_A_PWGT1 - R01_A_PWGT100 / adjfay=2.040816;
     model respvar = cov1 cov2;
     title 'SUDAAN proc regress using BRR-Fay Replication';
run;
```

The following code fits a logistic regression model using dichotomous variable *respvar2* (with values of 0 and 1) as the outcome variable and continuous variables *cov1* and *cov2* as the predictors. All parameter estimates are weighted (using the weight R01_A_PWGT) and the standard errors are calculated using the replicate weights (R01_A_PWGT1 - R01_A_PWGT100).

```
proc rlogist data=analysis_dataset filetype=sas design=brr;
     weight R01_A_PWGT;
     repwgt R01_A_PWGT1 - R01_A_PWGT100 / adjfay=2.040816;
     model respvar2 = cov1 cov2;
     title 'SUDAAN proc rlogist using BRR-Fay Replication';
run;
```

Domain analyses can be performed using the same examples above with the inclusion of a subpopn statement. For example, for frequencies *var1*, *var2*, *var3*, and *var4* for just those respondents with *domainvar* = 1, the following code could be used:

```
proc crosstab data=analysis_dataset filetype=sas design=brr;
     subpopn domainvar = 1;
     weight R01_A_PWGT;
     repwgt R01_A_PWGT1 - R01_A_PWGT100 / adjfay=2.040816;
     tables var1 var2 var3 var4;
     class var1 var2 var3 var4;
     print / style=nchs tablecell=all;
     title 'SUDAAN proc crosstab using BRR-Fay Replication';
run;
```

Stata

The full-sample weight, variance estimation method, BRR-Fay replicate weights, and Fay's factor are relayed to Stata using the `svyset` statement. The following statement should be used with the PATH Study data:

```
svyset [pweight= r01_a_pwgt], brr(r01_a_pwgt1 - r01_a_pwgt100)
vce(brr) mse fay(.3)
```

Assuming this `svyset` is used, the following code creates tables including the unweighted frequencies of categorical variables *var1*, *var2*, *var3*, and *var4*, weighted estimates of population totals and population proportions for each level of those variables (using the weight R01_A_PWGT) along with the standard errors of these estimates (using the replicate weights R01_A_PWGT1 - R01_A_PWGT100), and Wilson confidence intervals of estimated population proportions for a categorical variable *var1* (using the weight R01_A_PWGT and the replicate weights R01_A_PWGT1 – R01_A_PWGT100):

For weighted frequencies/estimates of population totals

```
svy: tabulate var1, count se
svy: tabulate var2, count se
svy: tabulate var3, count se
svy: tabulate var4, count se
```

For weighted frequencies/estimates of population proportions

```
svy: tabulate var1, se obs percent
svy: tabulate var2, se obs percent
svy: tabulate var3, se obs percent
svy: tabulate var4, se obs percent
```

Note that if all four variables are used in a single tabulate statement, the result is a multi-dimensional table containing all four variables rather than four one-dimensional tables.

To create these estimates for respondents with *domainvar* = 1, the following code may be used for each variable:

```
svy: tabulate var1, count se subpop(domainvar)
svy: tabulate var1, se obs percent
     subpop(domainvar)
```

Note that in the example above, Stata assumes that the subdomain of interest is *domainvar* = 1.

For Wilson confidence intervals of estimated population proportions

```
svy: proportion var1, citype(wilson)
```

To create Wilson confidence intervals of estimated population proportions for each level of *domainvar*, the following code may be used:

```
svy: proportion var1, over(domainvar) citype(wilson)
```

Assuming the `svyset` statement above is used, the following code creates the weighted mean of continuous variable *var5* (using the weight R01_A_PWGT) along with the standard error of that estimate (using the replicate weights R01_A_PWGT1 - R01_A_PWGT100):

```
svy: mean var5
```

To create these estimates for each of the levels of *domainvar*, the following code may be used:

```
svy: mean var5, over(domainvar)
```

Assuming the `svyset` statement above is used, the following code fits a linear regression model using continuous variable *respvar* as the outcome variable and continuous variables *cov1* and *cov2* as the predictors. All parameter estimates are weighted (using the weight R01_A_PWGT) and the standard errors are calculated using the replicate weights (R01_A_PWGT1 - R01_A_PWGT100):

```
svy: regress respvar cov1 cov2
```

To perform this regression for respondents with *domainvar* = 1, the following code may be used:

```
svy, subpop(if domainvar==1): regress respvar cov1 cov2
```

The following code fits a logistic regression model using dichotomous variable *respvar2* (with values of 0 and 1) as the outcome variable and continuous variables *cov1* and *cov2* as the

predictors. All parameter estimates are weighted (using the weight R01_A_PWGT) and the standard errors are calculated using the replicate weights (R01_A_PWGT1 - R01_A_PWGT100).

For odds ratios

```
svy: logistic respvar2 cov1 cov2
```

For coefficient estimates

```
svy: logit respvar2 cov1 cov2
```

To perform this regression for respondents with *domainvar* = 1, the following code may be used:

```
svy, subpop(if domainvar==1): logistic respvar2 cov1 cov2
svy, subpop(if domainvar==1): logit respvar2 cov1 cov2
```

R

The full-sample weight, variance estimation method, BRR-Fay replicate weights, and Fay's factor are relayed to the R survey package using the `svrepdesign` function. The following should be used with the PATH Study data:

```
options(survey.replicates.mse=TRUE)
y <-
     svrepdesign(
          id = ~PERSONID,
          weights = ~R01_A_PWGT,
          repweights = "R01_A_PWGT[1-9]+",
          type = "Fay",
          rho = 0.3,
          data = analysis_dataset
     )
```

Assuming the options and `svrepdesign` function above are executed, the following code creates the estimates of population totals and population proportions of the categorical variables *var1*, *var2*, *var3*, and *var4* (using the weight R01_A_PWGT) along with the standard errors of these estimates (using the replicate weights R01_A_PWGT1 - R01_A_PWGT100) :

For weighted frequencies/estimates of population totals

```
svytotal(~factor(var1), design=y, na.rm=T)
svytotal(~factor(var2), design=y, na.rm=T)
svytotal(~factor(var3), design=y, na.rm=T)
svytotal(~factor(var4), design=y, na.rm=T)
```

For weighted frequencies/estimates of population proportions

```
svymean(~factor(var1), design=y, na.rm=T)
svymean(~factor(var2), design=y, na.rm=T)
svymean(~factor(var3), design=y, na.rm=T)
svymean(~factor(var4), design=y, na.rm=T)
```

Note that this code should be executed for one variable at a time or the function will output estimates using only those records that have non-missing values across all the variables specified.

For confidence intervals

```
svyciprop(~var1, y, method = "lo", level = 0.95)
svyciprop(~var2, y, method = "lo", level = 0.95)
svyciprop(~var3, y, method = "lo", level = 0.95)
svyciprop(~var4, y, method = "lo", level = 0.95)
```

To create these estimates for each of the levels of *domainvar*, the following code may be used:

```
svyby(~factor(var1), ~domainvar, svytotal, design=y, na.rm=T)
svyby(~factor(var1), ~domainvar, svymean, design=y, na.rm=T)
```

Assuming the options and `svrepdesign` function above are executed, the following code creates the weighted mean of continuous variable *var5* (using the weight R01_A_PWGT) along with the appropriate standard errors of that estimate (using the replicate weights R01_A_PWGT1 - R01_A_PWGT100):

```
svymean(~var5, design = y,na.rm=T)
```

To create these estimates for each of the levels of *domainvar*, the following code may be used:

```
svyby(~var5,~domainvar, svymean, design=y, na.rm=T)
```

Assuming the options and `svrepdesign` function above are executed, the following code fits a linear regression model using continuous variable *respvar* as the outcome variable and continuous variables *cov1* and *cov2* as the predictors. All parameter estimates are weighted (using the weight R01_A_PWGT) and the standard errors are calculated using the replicate weights (R01_A_PWGT1 - R01_A_PWGT100).

For coefficient estimates

```
svyglm(respvar ~ cov1 + cov2, design=y)
```

For standard errors and significance test of model coefficients

```
summary(svyglm(respvar ~ cov1 + cov2, design=y))
```

The following code fits a logistic regression model using dichotomous variable *respvar2* (with values of 0 and 1) as the outcome variable and continuous variables *cov1* and *cov2* as the predictors. All parameter estimates are weighted (using the weight R01_A_PWGT) and the standard errors are calculated using the replicate weights (R01_A_PWGT1 - R01_A_PWGT100).

For coefficient estimates

```
svyglm(respvar2 ~ cov1 + cov2, design=y, family=binomial)
```

For standard errors and significance test of model coefficients

```
summary(svyglm(respvar2 ~ cov1 + cov2, design=y,
family=binomial))
```

To perform these regressions for respondents with *domainvar* = 1, the design statement should be changed as follows:

```
design=subset(y, domainvar=="1"))
```

This design statement applies to both linear and logistic regressions.

SPSS

SPSS does not have the functionality to create variance estimates using the replication method, the preferred method for the PATH Study. As such, only the Taylor series linearization approach to

variance estimation may be used with this software package. The appropriate design variables and full-sample weights are relayed to SPSS through a plan file created using the `csplan` function.

The following code should be used to create a plan file for use with the PATH Study data:

```
csplan analysis
/plan file="c:\myspace\myplan.csaplan"
/planvars analysisweight=R01_A_PWGT
/print plan
/design strata=VARSTRAT cluster=VARPSU
/estimator type=wr.
```

Assuming the plan file is created as specified above, the following code creates tables including the frequencies of categorical variables *var1*, *var2*, *var3*, and *var4*, and estimates of population totals and population proportions for each level of those variables (using the weight R01_A_PWGT) along with the standard errors of these estimates (created using linearization):

```
cstabulate
/plan file="c:\myspace\myplan.csaplan"
/tables variables=var1 var2 var3 var4
/cells popsize tablepct
/statistics count se cin(95).
```

Because SPSS only supports the Taylor series linearization approach to variance estimation, analyses for a subpopulation must be made using all cases rather than conducting the analysis only on a subset of cases of analytic interest. Estimates for each of the levels of *domainvar* may be created by including the following statement:

```
/subpop table=domainvar
```

Assuming the plan file is created as specified above, the following code creates the weighted mean of continuous variable *var5* (using the weight R01_A_PWGT) along with the appropriate standard error of that estimate (using linearization):

```
csdescriptives
/plan file="c:\myspace\myplan.csaplan"
/summary variables var5
/mean
/statistics se cin(95).
```

Estimates for each of the levels of *domainvar* may be created by including the following statement:

```
/subpop table=domainvar
```

Assuming the plan file is created as specified above, the following code fits a linear regression model using continuous variable *respvar* as the outcome variable and continuous variables *cov1* and *cov2* as the predictors. All parameter estimates are weighted (using the weight R01_A_PWGT) and the standard errors are calculated using linearization.

```
csglm respvar with cov1 cov2
/plan file="c:\myspace\myplan.csaplan"
/model cov1 cov2
/statistics parameter se cinterval.
```

The following code fits a logistic regression model using dichotomous variable *respvar2* (with values of 0 and 1) as the outcome variable and continuous variables *cov1* and *cov2* as the predictors. All parameter estimates are weighted (using the weight R01_A_PWGT) and the standard errors are calculated using linearization.

```
cslogistic respvar2 with cov1 cov2
/plan file="c:\myspace\myplan.csaplan"
/model cov1 cov2
/statistics parameter se cinterval.
```

To perform either of these regressions for respondents with *domainvar* = 1, the following statement may be inserted:

```
/domain variable domainvar (1)
```

Appendix B
Excluded Questionnaire Variables

Table B-1. Wave 1 questionnaire variables entirely excluded from the PUFs

Questionnaire variable	Instrument	Questionnaire variable description
R01_AC1033MC_OS	Adult	Retail location where your cigarette is purchased most of the time - Specify
R01_AC1033RY_OS	Adult	Retail location where your roll your own cigarette tobacco is purchased most of the time - Specify
R01_AC1049MC	Adult	Brand of cigarettes usually / last smoked - Specify
R01_AC1049RY	Adult	Brand of roll-your-own cigarette tobacco usually / last smoked - Specify
R01_AC1071MC	Adult	Sub-brand of cigarette product usually / last smoked - Specify
R01_AC1071RY	Adult	Name of roll-your-own cigarette tobacco product usually / last smoked - Specify
R01_AD1033_OS	Adult	Retail location where your dissolvable tobacco is purchased most of the time - Specify
R01_AD1049	Adult	Brand of dissolvable tobacco usually / last used - Specify
R01_AD1071	Adult	Sub-brand of dissolvable tobacco usually / last smoked - Specify
R01_AE1033_OS	Adult	Retail location where your e-cigarettes and e-liquid / e-cigarette cartridges is purchased most of the time - Specify
R01_AE1049	Adult	Brand of e-cigarettes usually / last used - Specify
R01_AE9003	Adult	Brand of e-cigarette owned - Specify
R01_AG1033CG_OS	Adult	Retail location where your cigarillos / cigarillos for blunts is purchased most of the time - Specify
R01_AG1033FC_OS	Adult	Retail location where your filtered cigars / filtered cigars for blunts is purchased most of the time - Specify
R01_AG1033TC_OS	Adult	Retail location where your traditional cigars are purchased most of the time - Specify
R01_AG1049CG	Adult	Brand of cigarillos usually / last smoked - Specify
R01_AG1049FC	Adult	Brand of filtered cigars usually / last smoked - Specify
R01_AG1049TC	Adult	Brand of traditional cigars usually / last smoked - Specify
R01_AG1071CG	Adult	Sub-brand of cigarillo product usually / last smoked - Specify
R01_AG1071FC	Adult	Sub-brand of filtered cigar product usually / last smoked - Specify
R01_AG1071TC	Adult	Sub-brand of traditional cigar product usually / last smoked - Specify
R01_AH1033_OS	Adult	Retail location where your hookah tobacco is purchased most of the time - Specify
R01_AH1049	Adult	Brand of hookah tobacco usually / last smoked - Specify
R01_AH1071	Adult	Sub-brand of hookah tobacco product usually / last smoked - Specify
R01_AH9011_OS	Adult	Place where usually smoke / smoked a hookah: Somewhere else - Specify
R01_AM0007	Adult	Ever served on active duty in the U.S. Armed Forces, Military Reserves or National Guard
R01_AM0008	Adult	Currently on active duty in the U.S. Armed Forces

Table B-1. Wave 1 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R01_AM0010_01	Adult	When served on active duty in the U.S. Armed Forces: September 2001 or later
R01_AM0010_02	Adult	When served on active duty in the U.S. Armed Forces: August 1990 to August 2001 (including Persian Gulf War)
R01_AM0010_03	Adult	When served on active duty in the U.S. Armed Forces: September 1980 to July 1990
R01_AM0010_04	Adult	When served on active duty in the U.S. Armed Forces: May 1975 to August 1980
R01_AM0010_05	Adult	When served on active duty in the U.S. Armed Forces: Vietnam era (August 1964 to April 1975)
R01_AM0010_06	Adult	When served on active duty in the U.S. Armed Forces: March 1961 to July 1964
R01_AM0010_07	Adult	When served on active duty in the U.S. Armed Forces: February 1955 to February 1961
R01_AM0010_08	Adult	When served on active duty in the U.S. Armed Forces: Korean War (July 1950 to January 1955)
R01_AM0010_09	Adult	When served on active duty in the U.S. Armed Forces: January 1947 to June 1950
R01_AM0010_10	Adult	When served on active duty in the U.S. Armed Forces: World War II (December 1941 to December 1946)
R01_AM0010_11	Adult	When served on active duty in the U.S. Armed Forces: November 1941 or earlier
R01_AM0011_01	Adult	Branch served when on active duty: Army
R01_AM0011_02	Adult	Branch served when on active duty: Navy
R01_AM0011_03	Adult	Branch served when on active duty: Air Force
R01_AM0011_04	Adult	Branch served when on active duty: Marine Corps
R01_AM0011_05	Adult	Branch served when on active duty: Coast Guard
R01_AM0012	Adult	Ever been enrolled in VA Health Care
R01_AM0013_01	Adult	Received any VA Health Care Benefits: I have received services at the VA
R01_AM0013_02	Adult	Received any VA Health Care Benefits: VA has paid for some or all of my health care
R01_AM0013_03	Adult	Received any VA Health Care Benefits: I have not received any VA health care benefits
R01_AM0016	Adult	Main job title or occupation
R01_AM0021	Adult	Sexual attraction to gender
R01_AM0025	Adult	Sexual orientation - Specify
R01_AM0050	Adult	Ever served on active duty in the U.S. Armed Forces, Military Reserves or National Guard (Confirmation)
R01_AP1033_OS	Adult	Retail location where your pipe tobacco is purchased most of the time - Specify
R01_AP1049	Adult	Brand of pipe tobacco usually / last smoked - Specify
R01_AP1071	Adult	Sub-brand of pipe tobacco product usually / last smoked - Specify
R01_AS1033SM_OS	Adult	Retail location where your smokeless tobacco is purchased most of the time - Specify
R01_AS1033SU_OS	Adult	Retail location where your snus pouches is purchased most of the time - Specify
R01_AS1049SM	Adult	Brand of smokeless tobacco usually / last used - Specify

Table B-1. Wave 1 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R01_AS1049SU	Adult	Brand of snus pouch usually / last used - Specify
R01_AS1071SM	Adult	Sub-brand of smokeless tobacco product usually / last smoked - Specify
R01_AS1071SU	Adult	Sub-brand of snus pouch product usually / last smoked - Specify
R01_AX0190	Adult	Is deaf or has serious difficulty hearing
R01_AX0191	Adult	Is blind or has serious difficulty seeing, even when wearing glasses
R01_AX0197_OS	Adult	Brand of favorite tobacco advertisement - Specify
R01_AX0217_10A	Adult	Tobacco Product 10, bar code scan
R01_AX0217_1A	Adult	Tobacco Product 1, bar code scan
R01_AX0217_2A	Adult	Tobacco Product 2, bar code scan
R01_AX0217_3A	Adult	Tobacco Product 3, bar code scan
R01_AX0217_4A	Adult	Tobacco Product 4, bar code scan
R01_AX0217_5A	Adult	Tobacco Product 5, bar code scan
R01_AX0217_6A	Adult	Tobacco Product 6, bar code scan
R01_AX0217_7A	Adult	Tobacco Product 7, bar code scan
R01_AX0217_8A	Adult	Tobacco Product 8, bar code scan
R01_AX0217_9A	Adult	Tobacco Product 9, bar code scan
R01_AX0309	Adult	Calendar year of most recent pregnancy
R01_LAD01	Adult	Respondent's preferred language to complete ACASI interview
R01_LCAD01	Adult	Language in which CAPI portions of adult interview were conducted
R01_CPT07	Parent	Youth date of birth
R01_LCPT01	Parent	Language in which the CAPI portions of the interview was completed (interviewer report)
R01_PM0016	Parent	Youth was identified as a twin or part of a multiple birth
R01_PM0017	Parent	Youth and sibling are identical twins
R01_PM0019	Parent	Any sibling in multiple birth identical to youth
R01_PM0021	Parent	First name of sibling that youth is a twin of
R01_PM0030	Parent	First names of siblings in multiple birth that are identical to youth
R01_PM0035	Parent	First names of siblings that are in the multiple birth with youth
R01_PT0001_OS_1	Parent	Parent or guardian relationship to youth: Other relative - specify
R01_PT0001_OS_2	Parent	Parent or guardian relationship to youth: Other non-relative - specify
R01_PT0002_OS_1	Parent	Parent's spouse or partner relationship to youth: Other relative - specify
R01_PT0002_OS_2	Parent	Parent's spouse or partner relationship to youth: Non-relative - specify
R01_PX0186	Parent	Youth has serious difficulty walking or climbing stairs
R01_PX0190	Parent	Youth is deaf or has serious difficulty hearing
R01_PX0191	Parent	Youth is blind or has serious difficulty seeing even when wearing glasses
R01_EYOUTH	Youth	Respondent is an emancipated youth
R01_LEY01	Youth	Emancipated Youth's preferred language to complete ACASI interview
R01_LYH01	Youth	Youth's preferred language to complete ACASI interview
R01_YB1118BD_OS	Youth	How you usually got your own bidis in the past 30 days - Specify

Table B-1. Wave 1 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R01_YB1033BD_OS	Youth	Retail location where your bidis are purchased most of the time - Specify
R01_YB1118KK_OS	Youth	How you usually got your own kreteks in the past 30 days - Specify
R01_YB1033KK_OS	Youth	Retail location where your kreteks are purchased most of the time - Specify
R01_YB1049KK	Youth	Brand of kreteks usually / last smoked - Specify
R01_YB1071KK	Youth	Sub-brand of kretek product usually/last smoked - Specify
R01_YC1118_OS	Youth	How you usually got your own cigarettes in the past 30 days - Specify
R01_YC1033_OS	Youth	Retail location where your cigarettes are purchased most of the time - Specify
R01_YC1049	Youth	Brand of cigarettes usually / last smoked - Specify
R01_YC1071	Youth	Sub-brand of cigarette product usually/last smoked - Specify
R01_YB1049BD	Youth	Brand of bidis usually / last smoked - Specify
R01_YD1118_OS	Youth	How you usually got your own dissolvable tobacco in the past 30 days - Specify
R01_YD1033_OS	Youth	Retail location where you usually buy dissolvable tobacco most of the time - Specify
R01_YD1049	Youth	Brand of dissolvable tobacco usually / last used - Specify
R01_YD1071	Youth	Sub-brand of dissolvable tobacco product usually / last smoked - Specify
R01_YE1118_OS	Youth	How you usually got your own e-cigarettes / cartridges and e-liquid in the past 30 days - Specify
R01_YE1033_OS	Youth	Retail location where your e-cigarettes / cartridges and e-liquid are purchased most of the time - Specify
R01_YE1049	Youth	Brand of e-cigarettes usually / last used - Specify
R01_YG1118CL_OS	Youth	How you usually got your own cigarillos in the past 30 days - Specify
R01_YG1033CL_OS	Youth	Retail location where your cigarillos are purchased most of the time - Specify
R01_YG1118FC_OS	Youth	How you usually got your own filtered cigars in the past 30 days - Specify
R01_YG1033FC_OS	Youth	Retail location where your filtered cigars are purchased most of the time - Specify
R01_YG1118TC_OS	Youth	How you usually got your own traditional cigars in the past 30 days - Specify
R01_YG1033TC_OS	Youth	Retail location where your traditional cigars are purchased most of the time - Specify
R01_YG1049CL	Youth	Brand of cigarillos usually / last smoked - Specify
R01_YG1049FC	Youth	Brand of filtered cigars usually / last smoked - Specify
R01_YG1049TC	Youth	Brand of traditional cigars usually / last smoked - Specify
R01_YG1071CL	Youth	Sub-brand of cigarillo product usually/last smoked - Specify
R01_YG1071FC	Youth	Sub-brand of filtered cigar product usually/last smoked - Specify
R01_YG1071TC	Youth	Sub-brand of traditional cigar product usually / last smoked - Specify
R01_YH9011_OS	Youth	Usually smoke a hookah: Somewhere else - Specify
R01_YH1118_OS	Youth	How you usually got your own shisha or hookah tobacco in the past 30 days - Specify

Table B-1. Wave 1 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R01_YH1033_OS	Youth	Retail location where your hookah tobacco is purchased most of the time - Specify
R01_YH1049	Youth	Brand of hookah tobacco usually / last smoked - Specify
R01_YH1071	Youth	Sub-brand of hookah tobacco product usually / last smoked - Specify
R01_YM0020	Youth	Last grade / year in school completed
R01_YM0021	Youth	Sexual attraction to gender
R01_YM0022	Youth	Sexual orientation (initial prompt)
R01_YM0023	Youth	Sexual orientation - other description
R01_YM0028	Youth	Sexual orientation (second prompt)
R01_YM0024	Youth	Reason for not providing sexual orientation
R01_YM0025	Youth	Sexual orientation - Specify
R01_YP1118_OS	Youth	How you usually got your own pipe tobacco in the past 30 days - Specify
R01_YP1033_OS	Youth	Retail location where your pipe tobacco are purchased most of the time - Specify
R01_YP1049	Youth	Brand of pipe tobacco usually / last smoked - Specify
R01_YP1071	Youth	Sub-brand of pipe tobacco product usually / last smoked - Specify
R01_YS1118SM_OS	Youth	How you usually got your own smokeless tobacco in the past 30 days - Specify
R01_YS1033SM_OS	Youth	Retail location where your smokeless tobacco is purchased most of the time - Specify
R01_YS1118SU_OS	Youth	How you usually got your own snus pouches in the past 30 days - Specify
R01_YS1033SU_OS	Youth	Retail location where your snus pouches are purchased most of the time - Specify
R01_YS1049SM	Youth	Brand of smokeless tobacco usually / last used - Specify
R01_YS1049SU	Youth	Brand of snus pouches usually / last used - Specify
R01_YS1071SM	Youth	Sub-brand of smokeless tobacco product usually / last smoked - Specify
R01_YS1071SU	Youth	Sub-brand of snus pouch product usually / last smoked - Specify
R01_YX0136	Youth	Currently pregnant
R01_YX0137_NN	Youth	Number of weeks / months pregnant - Number
R01_YX0137_UN	Youth	Number of weeks / months pregnant - Unit
R01_YX0186	Youth	Has serious difficulty walking or climbing stairs
R01_YX0190	Youth	Is deaf or has serious difficulty hearing
R01_YX0191	Youth	Is blind or has serious difficulty seeing, even when wearing glasses
R01_YX0197_OS	Youth	Brand of favorite tobacco advertisement - Specify
R01_YX0478_OS	Youth	Received discount coupons from: Some other way - Specify
R01_YX0480_OS	Youth	How tobacco company sent you the information other than coupons: Some other way - Specify
R01_YY0601_OS	Youth	First type of tobacco you tried - Specify

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_AM0001 R01_AM0002 R01_AM0003	Adult	R01R_A_AGECAT7	Age range when interviewed (7 levels)
R01_AM0004	Adult	R01R_A_SEX	Adult respondent sex
R01_AM0005_01 R01_AM0005_02 R01_AM0005_03 R01_AM0005_04 R01_AM0005_05	Adult	R01R_A_HISP	Hispanic origin from the interview (2 levels)
R01_AM0006_WH R01_AM0006_AA R01_AM0006_AI R01_AM0006_IN R01_AM0006_CH R01_AM0006_FI R01_AM0006_JA R01_AM0006_KO R01_AM0006_VI R01_AM0006_OA R01_AM0006_HI R01_AM0006_GU R01_AM0006_SA R01_AM0006_OP	Adult	R01R_A_RACECAT3	Race from the interview (3 levels)
R01_AM0018	Adult	R01R_A_AM0018	Highest grade or level of school completed (6 levels)
R01_AM0030	Adult	R01R_A_AM0030	Total household income in the past 12 months (5 levels)
R01_AX0135_01 R01_AX0135_02 R01_AX0135_03 R01_AX0135_04 R01_AX0135_05	Adult	R01R_A_AX0135	Ever had pregnancy outcome of miscarriages, induced abortions, ectopic or tubal pregnancies or stillbirths (2 levels)
R01_AX0300_01 R01_AX0300_02 R01_AX0300_03 R01_AX0300_04 R01_AX0300_05 R01_AX0300_06 R01_AX0300_07	Adult	R01R_A_AX0300	Indicator of Ever had outcome for live birth of Preterm birth, low birth weight, birth defects, Placenta Previa, Placenta Abruption, or Pre-eclampsia (2 levels)
R01_AC1006 R01_AC1120	Adult	R01R_A_AC1006	Age range when first smoked part or all of a cigarette (6 levels)
R01_AC1007 R01_AC1121	Adult	R01R_A_AC1007	Age range when first started smoking cigarettes fairly regularly (6 levels)
R01_AC1020 R01_AC1122	Adult	R01R_A_AC1020	Age range when first started smoking cigarettes every day (6 levels)
R01_AD1006 R01_AD1120	Adult	R01R_A_AD1006	Age range when first used dissolvable tobacco, even one or two times (6 levels)

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_AD1007 R01_AD1121	Adult	R01R_A_AD1007	Age range when first started using dissolvable tobacco fairly regularly (6 levels)
R01_AD1020 R01_AD1122	Adult	R01R_A_AD1020	Age range when first started using dissolvable tobacco every day (6 levels)
R01_AE1006 R01_AE1120	Adult	R01R_A_AE1006	Age range when first time used an e-cigarette, even one or two times (6 levels)
R01_AE1007 R01_AE1121	Adult	R01R_A_AE1007	Age range when first started using e-cigarettes fairly regularly (6 levels)
R01_AE1020 R01_AE1122	Adult	R01R_A_AE1020	Age range when first started using e-cigarettes every day (6 levels)
R01_AG1006CG R01_AG1120CG	Adult	R01R_A_AG1006CG	Age range when first time smoked part or all of a cigarillo, even one or two puffs (6 levels)
R01_AG1006FC R01_AG1120FC	Adult	R01R_A_AG1006FC	Age range when first smoked part or all of a filtered cigar, even one or two puffs (6 levels)
R01_AG1006TC R01_AG1120TC	Adult	R01R_A_AG1006TC	Age range when first smoked part or all of a traditional cigar, even one or two puffs (6 levels)
R01_AG1007CG R01_AG1121CG	Adult	R01R_A_AG1007CG	Age range when first started smoking cigarillos fairly regularly (6 levels)
R01_AG1007FC R01_AG1121FC	Adult	R01R_A_AG1007FC	Age range when first started smoking filtered cigars fairly regularly (6 levels)
R01_AG1007TC R01_AG1121TC	Adult	R01R_A_AG1007TC	Age range when first started smoking traditional cigars fairly regularly (6 levels)
R01_AG1020CG R01_AG1122CG	Adult	R01R_A_AG1020CG	Age range when first started smoking cigarillos every day (6 levels)
R01_AG1020FC R01_AG1122FC	Adult	R01R_A_AG1020FC	Age range when first started smoking filtered cigars every day (6 levels)
R01_AG1020TC R01_AG1122TC	Adult	R01R_A_AG1020TC	Age range when first started smoking traditional cigars every day (6 levels)
R01_AH1006 R01_AH1120	Adult	R01R_A_AH1006	Age range when first smoked hookah, even one or two puffs (6 levels)
R01_AH1007 R01_AH1121	Adult	R01R_A_AH1007	Age range when first started smoking hookah fairly regularly (6 levels)
R01_AH1020 R01_AH1122	Adult	R01R_A_AH1020	Age range when first started smoking hookah every day (6 levels)

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_AP1006 R01_AP1120	Adult	R01R_A_AP1006	Age range when first smoked part or all of a pipe filled with tobacco, even one or two puffs (6 levels)
R01_AP1007 R01_AP1121	Adult	R01R_A_AP1007	Age range when first started smoking a pipe filled with tobacco fairly regularly (6 levels)
R01_AP1020 R01_AP1122	Adult	R01R_A_AP1020	Age range when first started smoking a pipe filled with tobacco every day (6 levels)
R01_AS1006SM R01_AS1120SM	Adult	R01R_A_AS1006SM	Age range when first used smokeless tobacco, even one or two times (6 levels)
R01_AS1006SU R01_AS1120SU	Adult	R01R_A_AS1006SU	Age range when first used snus pouches, even one or two times (6 levels)
R01_AS1007SM R01_AS1121SM	Adult	R01R_A_AS1007SM	Age range when first started using smokeless tobacco fairly regularly (6 levels)
R01_AS1007SU R01_AS1121SU	Adult	R01R_A_AS1007SU	Age range when first started using snus pouches fairly regularly (6 levels)
R01_AS1020SM R01_AS1122SM	Adult	R01R_A_AS1020SM	Age range when first started using smokeless tobacco every day (6 levels)
R01_AS1020SU R01_AS1122SU	Adult	R01R_A_AS1020SU	Age range when first started using snus pouches every day (6 levels)
R01_AX0074 R01_AX0270	Adult	R01R_A_AX0074	Age range when first alcoholic drink was consumed (6 levels)
R01_AX0079 R01_AX0271	Adult	R01R_A_AX0079	Age range when first used marijuana, hash, THC or grass (6 levels)
R01_AX0082_01 R01_AX0272_01	Adult	R01R_A_AX0082_01	Age range when first started using: Ritalin or Adderall (6 levels)
R01_AX0082_02 R01_AX0272_02	Adult	R01R_A_AX0082_02	Age range when first started using: Painkillers, sedatives or tranquilizers (6 levels)
R01_AX0082_03 R01_AX0272_03	Adult	R01R_A_AX0082_03	Age range when first started using: Cocaine or crack (6 levels)
R01_AX0082_04 R01_AX0272_04	Adult	R01R_A_AX0082_04	Age range when first started using: Stimulants like methamphetamine or speed (6 levels)
R01_AX0082_05 R01_AX0272_05	Adult	R01R_A_AX0082_05	Age range when first started using: Any other drugs like heroin, inhalants, solvents, or hallucinogens (6 levels)
R01_AX0086 R01_AX0087	Adult	R01R_A_AX0086	Age range when first drank alcohol at all, counting small tastes or sips (6 levels)

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_AX0112 R01_AX0252	Adult	R01R_A_AX0112	Age range when you were first told you had a heart attack (6 levels)
R01_AX0114 R01_AX0253	Adult	R01R_A_AX0114	Age range when you were first told you had high blood pressure (6 levels)
R01_AX0115 R01_AX0254	Adult	R01R_A_AX0115	Age range when you were first told you had high cholesterol (6 levels)
R01_AX0116 R01_AX0255	Adult	R01R_A_AX0116	Age range when you were first told you had congestive heart failure (6 levels)
R01_AX0117 R01_AX0256	Adult	R01R_A_AX0117	Age range when you were first told you had a stroke (6 levels)
R01_AX0120 R01_AX0257	Adult	R01R_A_AX0120	Age range when you were first told you had COPD (6 levels)
R01_AX0121 R01_AX0258	Adult	R01R_A_AX0121	Age range when you were first told you had chronic bronchitis (6 levels)
R01_AX0123 R01_AX0259	Adult	R01R_A_AX0123	Age range when you were first told you had emphysema (6 levels)
R01_AX0124 R01_AX0260	Adult	R01R_A_AX0124	Age range when you were first told you had asthma (6 levels)
R01_AX0131 R01_AX0261	Adult	R01R_A_AX0131	Age range when you were first told you had gum disease (6 levels)
R01_AX0133 R01_AX0262	Adult	R01R_A_AX0133	Age range when you were first told you had pre-cancerous oral lesions (6 levels)
R01_AX0143 R01_AX0264	Adult	R01R_A_AX0143	Age range when you were first told you had an ulcer (6 levels)
R01_AX0148 R01_AX0266	Adult	R01R_A_AX0148	Age range when you were first told you had stomach or gastro-intestinal bleeding (6 levels)
R01_AX0150 R01_AX0267	Adult	R01R_A_AX0150	Age range when you were first told you had osteoporosis (6 levels)
R01_AX0152 R01_AX0269	Adult	R01R_A_AX0152	Age range when you were first told you had a cataract or glaucoma (6 levels)
R01_AX0198 R01_AX0268	Adult	R01R_A_AX0198	Age range when you were first told you had a bone fracture because you have fragile bones (6 levels)
R01_AX0280 R01_AX0263	Adult	R01R_A_AX0280	Age range when you were first told you had diabetes, sugar diabetes, high blood sugar or borderline diabetes (6 levels)

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_AM0026_01 R01_AM0026_02 R01_AM0026_03 R01_AM0026_04 R01_AM0026_05 R01_AM0026_06 R01_AM0026_07 R01_AM0026_08	Adult	R01R_A_AM0026	Currently covered by health insurance or health coverage plan (5 levels)
R01_AX0308	Adult	R01R_A_AX0308	Number of pregnancies that have resulted in a live birth (2 levels)
R01_AX0310 R01_AX0108_FT R01_AX0108_IN R01_AX0311 R01_AX0109 R01_AX0312	Adult	R01R_A_BMI	Body mass index
R01_AX0145_01 R01_AX0145_06 R01_AX0145_07 R01_AX0145_08 R01_AX0145_10 R01_AX0145_11 R01_AX0145_13 R01_AX0145_14 R01_AX0145_17 R01_AX0145_20 R01_AX0145_22 R01_AX0145_26 R01_AX0145_28	Adult	R01R_A_AX0145_TOB	Type of cancer is a tobacco related cancer

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_AX0146_01 R01_AX0146_06 R01_AX0146_07 R01_AX0146_08 R01_AX0146_10 R01_AX0146_11 R01_AX0146_13 R01_AX0146_14 R01_AX0146_17 R01_AX0146_20 R01_AX0146_22 R01_AX0146_26 R01_AX0146_28 R01_AX0265_01 R01_AX0265_06 R01_AX0265_07 R01_AX0265_08 R01_AX0265_10 R01_AX0265_11 R01_AX0265_13 R01_AX0265_14 R01_AX0265_17 R01_AX0265_20 R01_AX0265_22 R01_AX0265_26 R01_AX0265_28	Adult	R01R_A_AX0146_TOB	Age when first tobacco-related cancer was diagnosed
R01_AX0145_02 R01_AX0145_03 R01_AX0145_04 R01_AX0145_05 R01_AX0145_09 R01_AX0145_12 R01_AX0145_15 R01_AX0145_16 R01_AX0145_18 R01_AX0145_19 R01_AX0145_21 R01_AX0145_23 R01_AX0145_24 R01_AX0145_25 R01_AX0145_27 R01_AX0145_29 R01_AX0145_30 R01_AX0145_31	Adult	R01R_A_AX0145_NONTOB	Type of cancer is a non-tobacco related cancer

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_AX0146_02 R01_AX0146_03 R01_AX0146_04 R01_AX0146_05 R01_AX0146_09 R01_AX0146_12 R01_AX0146_15 R01_AX0146_16 R01_AX0146_18 R01_AX0146_19 R01_AX0146_21 R01_AX0146_23 R01_AX0146_24 R01_AX0146_25 R01_AX0146_27 R01_AX0146_29 R01_AX0146_30 R01_AX0146_31 R01_AX0265_02 R01_AX0265_03 R01_AX0265_04 R01_AX0265_05 R01_AX0265_09 R01_AX0265_12 R01_AX0265_15 R01_AX0265_16 R01_AX0265_18 R01_AX0265_19 R01_AX0265_21 R01_AX0265_23 R01_AX0265_24 R01_AX0265_25 R01_AX0265_27 R01_AX0265_29 R01_AX0265_30 R01_AX0265_31	Adult	R01R_A_AX0146_NONTOB	Age when first non-tobacco-related cancer was diagnosed
R01_AM0022 R01_AM0023 R01_AM0024 R01_AM0028	Adult	R01R_A_SEXORIENT2	Adult sexual orientation (2 levels)
R01_PT0001	Parent	R01R_Y_PT0001	Parent or guardian relationship to youth (4 levels)
R01_PT0002	Parent	R01R_Y_PT0002	Parent's spouse or partner relationship to youth (4 levels)
R01_PM0001	Parent	R01R_Y_PM0001	Highest grade or year of school completed by parent (5 levels)
R01_PT0047	Parent	R01R_Y_PT0047	Parent or guardian: marital status (3 levels)

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_CPT07	Parent	R01R_Y_AGECAT2	Age range when interviewed (2 levels)
R01_YM0004	Youth	R01R_Y_SEX	Youth respondent sex
R01_YM0005_01 R01_YM0005_02 R01_YM0005_03 R01_YM0005_04 R01_YM0005_05	Youth	R01R_Y_HISP	Hispanic origin from the interview (2 levels)
R01_YM0006_01 R01_YM0006_02 R01_YM0006_03 R01_YM0006_04 R01_YM0006_05 R01_YM0006_06 R01_YM0006_07 R01_YM0006_08 R01_YM0006_09 R01_YM0006_10 R01_YM0006_11 R01_YM0006_12 R01_YM0006_13 R01_YM0006_14	Youth	R01R_Y_RACECAT3	Race from the interview (3 levels)
R01_YX0671_01 R01_YX0671_02 R01_YX0671_03 R01_YX0671_04 R01_YX0671_05	Youth	R01R_Y_YX0671	Anyone who lives with you now use tobacco (3 levels)
R01_YX0086 R01_YX0087	Youth	R01R_Y_YX0086	Age range when first drank alcohol at all, counting small tastes or sips (3 levels)
R01_YX0074 R01_YX0270	Youth	R01R_Y_YX0074	Age range when had first alcoholic drink, other than small tastes or sips (3 levels)
R01_YX0079 R01_YX0271	Youth	R01R_Y_YX0079	Age range when first used marijuana, hash, THC or grass (3 levels)
R01_YE1006 R01_YE1120	Youth	R01R_Y_YE1006	Age range when first tried an e-cigarette, even one or two times (3 levels)
R01_YP1006 R01_YP1120	Youth	R01R_Y_YP1006	Age range when first tried pipe tobacco, even one or two puffs (3 levels)
R01_YH1006 R01_YH1120	Youth	R01R_Y_YH1006	Age range when first tried smoking a hookah, even one or two puffs (3 levels)
R01_YB1006BD R01_YB1120BD	Youth	R01R_Y_YB1006BD	Age range when first tried a bidi, even one or two puffs (3 levels)
R01_YB1006KK R01_YB1120KK	Youth	R01R_Y_YB1006KK	Age range when first tried a kretek, even one or two puffs (3 levels)

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_YC1006 R01_YC1120	Youth	R01R_Y_YC1006	Age range when first tried cigarette smoking, even one or two puffs (3 levels)
R01_YD1006 R01_YD1120	Youth	R01R_Y_YD1006	Age range when first tried a dissolvable tobacco product, even one or two times (3 levels)
R01_YG1006CL R01_YG1120CL	Youth	R01R_Y_YG1006CL	Age range when first tried a cigarillo, even one or two puffs (3 levels)
R01_YG1006FC R01_YG1120FC	Youth	R01R_Y_YG1006FC	Age range when first tried a filtered cigar, even one or two puffs (3 levels)
R01_YG1006TC R01_YG1120TC	Youth	R01R_Y_YG1006TC	Age range when first tried a traditional cigar, even one or two puffs (3 levels)
R01_YS1006SM R01_YS1120SM	Youth	R01R_Y_YS1006SM	Age range when first tried smokeless tobacco, even one or two times (3 levels)
R01_YS1006SU R01_YS1120SU	Youth	R01R_Y_YS1006SU	Age range when first tried snus pouches, even one or two times (3 levels)
R01_YX0082_01 R01_YX0272_01	Youth	R01R_Y_YX0082_01	Age range when first used: Ritalin or Adderall (3 levels)
R01_YX0082_02 R01_YX0272_02	Youth	R01R_Y_YX0082_02	Age range when first used: Painkillers, sedatives or tranquilizers (3 levels)
R01_YX0082_03 R01_YX0272_03	Youth	R01R_Y_YX0082_03	Age range when first used: Cocaine or crack (3 levels)
R01_YX0082_04 R01_YX0272_04	Youth	R01R_Y_YX0082_04	Age range when first used: Stimulants like methamphetamine or speed (3 levels)
R01_YX0082_05 R01_YX0272_05	Youth	R01R_Y_YX0082_05	Age range when first used: Any other drugs like heroin, inhalants, solvents or hallucinogens (3 levels)
R01_PT0019 R01_YT0019	Youth/ Parent	R01R_Y_PY0019	Youth's grade performance in school in past 12 months (as reported by parent or emancipated youth)
R01_PT0022 R01_YT0022	Youth/ Parent	R01R_Y_PY0022	Youth has taken medications regularly for asthma in past 12 months (as reported by parent or emancipated youth)
R01_PT0030 R01_YT0030	Youth/ Parent	R01R_Y_PY0030	How often youth missed school due to illness in past 12 months (as reported by parent or emancipated youth)

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_PT0031 R01_YT0031	Youth/ Parent	R01R_Y_PY0031	Child has ever been told by a doctor or other health professional that he / she has asthma (as reported by parent or emancipated youth)
R01_PT0033 R01_YT0033	Youth/ Parent	R01R_Y_PY0033	Youth has been told by a doctor or other health professional in past 12 months that he / she has bronchitis, pneumonia or chronic cough (as reported by parent or emancipated youth)
R01_PT0034 R01_YT0034	Youth/ Parent	R01R_Y_PY0034	Number of visits youth made to ER or urgent care for a health problem in past 12 months (as reported by parent or emancipated youth)
R01_PT0035 R01_YT0035	Youth/ Parent	R01R_Y_PY0035	Youth's overall health status (as reported by parent or emancipated youth)
R01_PT0036_01 R01_YT0036_01	Youth/ Parent	R01R_Y_PY0036_01	Medications youth has taken for asthma in past 12 months: Quick-relief inhaler (as reported by parent or emancipated youth)
R01_PT0036_02 R01_YT0036_02	Youth/ Parent	R01R_Y_PY0036_02	Medications youth has taken for asthma in past 12 months: Controller / long-acting inhaler (as reported by parent or emancipated youth)
R01_PT0036_03 R01_YT0036_03	Youth/ Parent	R01R_Y_PY0036_03	Medications youth has taken for asthma in past 12 months: Other controlling medication (as reported by parent or emancipated youth)
R01_PT0036_04 R01_YT0036_04	Youth/ Parent	R01R_Y_PY0036_04	Medications youth has taken for asthma in past 12 months: Oral or injected steroid medication (as reported by parent or emancipated youth)
R01_PT0036_05 R01_YT0036_05	Youth/ Parent	R01R_Y_PY0036_05	Medications youth has taken for asthma in past 12 months: Oxygen therapy (as reported by parent or emancipated youth)
R01_PT0036_06 R01_YT0036_06	Youth/ Parent	R01R_Y_PY0036_06	Medications youth has taken for asthma in past 12 months: Other asthma medication (as reported by parent or emancipated youth)
R01_PT0039 R01_YT0039	Youth/ Parent	R01R_Y_PY0039	Number of asthma attacks youth has had in past 12 months that required use of an oral or injected steroid medication (as reported by parent or emancipated youth)

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_PT0040 R01_YT0040	Youth/ Parent	R01R_Y_PY0040	Youth has ever been told by a doctor or other health professional that he/she has bronchitis, pneumonia or chronic cough (as reported by parent or emancipated youth)
R01_PT0044 R01_YT0044	Youth/ Parent	R01R_Y_PY0044	Youth has ever been told by a doctor, dentist or other health professional that he/she has dental health issues (as reported by parent or emancipated youth)
R01_PT0048 R01_YT0048	Youth/ Parent	R01R_Y_PY0048	Youth has taken medications regularly for ADHD or ADD in past 12 months (as reported by parent or emancipated youth)
R01_PT0049 R01_YT0049	Youth/ Parent	R01R_Y_PY0049	Youth has been told by a doctor, dentist or other health professional in past 12 months that he/she has dental health issues (as reported by parent or emancipated youth)
R01_PT0050 R01_YT0050	Youth/ Parent	R01R_Y_PY0050	Youth has ever been told by a doctor or other health professional that he/she has high blood pressure (as reported by parent or emancipated youth)
R01_PT0051 R01_YT0051	Youth/ Parent	R01R_Y_PY0051	Youth has ever been told by a doctor or other health professional that he / she has high cholesterol (as reported by parent or emancipated youth)
R01_PT0052 R01_YT0052	Youth/ Parent	R01R_Y_PY0052	Youth has ever been told by a doctor or other health professional that he/she has ADHD or ADD (as reported by parent or emancipated youth response)
R01_PT0125 R01_YT0125	Youth/ Parent	R01R_Y_PY0125	Youth had an asthma attack in past 12 months that required use of an oral or injected steroid medication (as reported by parent or emancipated youth)
R01_PT0127 R01_YT0127	Youth/ Parent	R01R_Y_PY0127	Youth has visited an emergency room or urgent care center in past 12 months because of asthma (as reported by parent or emancipated youth)
R01_PX0187 R01_YX0187	Youth/ Parent	R01R_Y_PY0187	Youth has difficulty doing errands alone because of a physical, mental or emotional condition (as reported by parent or emancipated youth0

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_PX0188 R01_YX0188	Youth/ Parent	R01R_Y_PY0188	Youth has difficulty dressing or bathing (as reported by parent or emancipated youth)
R01_PX0189 R01_YX0189	Youth/ Parent	R01R_Y_PY0189	Youth has serious difficulty concentrating, remembering or making decisions because of a physical, mental or emotional condition (as reported by parent or emancipated youth)
R01_PT0281 R01_YT0281	Youth/ Parent	R01R_Y_PY0281	Youth has ever been told by a doctor or other health professional that he/she has diabetes, sugar diabetes, high blood sugar or borderline diabetes (as reported by parent or emancipated youth)
R01_PT0282 R01_YT0282	Youth/ Parent	R01R_Y_PY0282	Number of times youth has visited an emergency room or urgent care center in past 12 months because of asthma (as reported by parent or emancipated youth)
R01_PX0302 R01_YX0302	Youth/ Parent	R01R_Y_PY0302	Youth visited ER or urgent care for a health problem in past 12 months (as reported by parent or emancipated youth)
R01_PT0038 R01_PT0260 R01_YT0038 R01_YT0260	Youth/ Parent	R01R_Y_PY0038	Age youth was first told he/she has asthma (as reported by parent or emancipated youth)
R01_PT0041 R01_PT0253 R01_YT0041 R01_YT0253	Youth/ Parent	R01R_Y_PY0041	Age youth was first told he/she had high blood pressure (as reported by parent or emancipated youth)
R01_PT0042 R01_PT0263 R01_YT0042 R01_YT0263	Youth/ Parent	R01R_Y_PY0042	Age youth was first told by a doctor or other health professional that he/she has diabetes, sugar diabetes, high blood sugar or borderline diabetes (as reported by parent or emancipated youth)
R01_PT0043 R01_PT0254 R01_YT0043 R01_YT0254	Youth/ Parent	R01R_Y_PY0043	Age youth was first told he/she has high cholesterol (as reported by parent or emancipated youth)

Table B-2. Wave 1 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R01_PT0007_FT R01_PT0007_IN R01_PT0007_MT R01_PT0008_LB R01_PT0008_KG R01_YT0007_FT R01_YT0007_IN R01_YX0310 R01_YX0311 R01_YT0008 R01_YX0312	Youth/ Parent	R01R_Y_BMI	Body mass index
R01_YM0018 R01_YM0019	Youth	R01R_Y_YM0018	Grade level (If on holiday or break - grade level entering when returning to school) (7 levels)

Table B-3. Wave 2 questionnaire variables entirely excluded from the PUFs

Questionnaire variable	Instrument	Questionnaire variable description
R02_CAD10	Adult	Confirm respondent DOB
R02_LAD01	Adult	Respondent's preferred language to complete ACASI interview
R02_AM0007	Adult	Ever served on active duty in the U.S. Armed Forces, Military Reserves or National Guard
R02_AM0011_01	Adult	Branch served when on active duty: Army
R02_AM0011_02	Adult	Branch served when on active duty: Navy
R02_AM0011_03	Adult	Branch served when on active duty: Air Force
R02_AM0011_04	Adult	Branch served when on active duty: Marine Corps
R02_AM0011_05	Adult	Branch served when on active duty: Coast Guard
R02_AO9035_OS	Adult	Ever used the following electronic nicotine product: Something else - specify
R02_AC1033RY_OS	Adult	Retail location where your roll-your-own cigarette tobacco is purchased most of the time - Specify
R02_AC1049RY	Adult	Brand of roll-your-own cigarette tobacco usually/last smoked - Specify
R02_AC1071RY	Adult	Sub-brand of roll-your-own cigarette tobacco product usually/last smoked - Specify
R02_AC1033MC_OS	Adult	Retail location where your cigarette is purchased most of the time - Specify
R02_AC1049MC	Adult	Brand of cigarettes usually/last smoked - Specify
R02_AC1071MC	Adult	Sub-brand of cigarette product usually/last smoked - Specify
R02_AE9003	Adult	Brand of e-cigarette owned - Specify
R02_AE1033_OS	Adult	Retail location where your e-cigarettes/e-cigarette cartridges/e-liquid is purchased most of the time - Specify
R02_AE1049	Adult	Brand of e-cigarettes usually/last used - Specify
R02_AG1033TC_OS	Adult	Retail location where your traditional cigars is purchased most of the time - Specify
R02_AG1049TC	Adult	Brand of traditional cigars usually/last smoked - Specify
R02_AG1071TC	Adult	Sub-brand of traditional cigar product usually/last smoked - Specify
R02_AG1033CG_OS	Adult	Retail location where your cigarillos is purchased most of the time - Specify
R02_AG1049CG	Adult	Brand of cigarillos usually/last smoked - Specify
R02_AG1071CG	Adult	Sub-brand of cigarillo product usually/last smoked - Specify
R02_AG1033FC_OS	Adult	Retail location where your filtered cigars is purchased most of the time - Specify
R02_AG1049FC	Adult	Brand of filtered cigars usually/last smoked - Specify
R02_AG1071FC	Adult	Sub-brand of filtered cigar product usually/last smoked - Specify
R02_AG1033TJ_OS	Adult	Retail location where your traditional cigars for blunts is purchased most of the time - Specify
R02_AG1049TJ	Adult	Brand of traditional cigars for blunts usually/last smoked - Specify
R02_AG1071TJ	Adult	Sub-brand of traditional cigar as a blunt product usually/last smoked - Specify

Table B-3. Wave 2 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R02_AG1033GJ_OS	Adult	Retail location where your cigarillos for blunts is purchased most of the time - Specify
R02_AG1049GJ	Adult	Brand of cigarillos for blunts usually/last smoked - Specify
R02_AG1071GJ	Adult	Sub-brand of cigarillo as a blunt product usually/last smoked - Specify
R02_AG1033FJ_OS	Adult	Retail location where your filtered cigars for blunts is purchased most of the time - Specify
R02_AG1049FJ	Adult	Brand of filtered cigars for blunts usually/last smoked - Specify
R02_AG1071FJ	Adult	Sub-brand of filtered cigar as a blunt product usually/last smoked - Specify
R02_AP1033_OS	Adult	Retail location where your pipe tobacco is purchased most of the time - Specify
R02_AP1049	Adult	Brand of pipe tobacco usually/last smoked - Specify
R02_AP1071	Adult	Sub-brand of pipe tobacco product usually/last smoked - Specify
R02_AH9011_OS	Adult	Place where usually smoke/smoked hookah: Somewhere else - Specify
R02_AH1033_OS	Adult	Retail location where your shisha or hookah tobacco is purchased most of the time - Specify
R02_AH1049	Adult	Brand of shisha or hookah tobacco usually/last smoked - Specify
R02_AH1071	Adult	Sub-brand of hookah tobacco product usually/last smoked - Specify
R02_AS1033SU_OS	Adult	Retail location where your snus pouches is purchased most of the time - Specify
R02_AS1049SU	Adult	Brand of snus pouch usually/last used - Specify
R02_AS1071SU	Adult	Sub-brand of snus pouch product usually/last used - Specify
R02_AS1033SM_OS	Adult	Retail location where your smokeless tobacco is purchased most of the time - Specify
R02_AS1049SM	Adult	Brand of smokeless tobacco usually/last used - Specify
R02_AS1071SM	Adult	Sub-brand of smokeless tobacco product usually/last used - Specify
R02_AD1033_OS	Adult	Retail location where your dissolvable tobacco is purchased most of the time - Specify
R02_AD1049	Adult	Brand of dissolvable tobacco usually/last used - Specify
R02_AD1071	Adult	Sub-brand of dissolvable tobacco usually/last smoked - Specify
R02_AX0190_NB	Adult	Is deaf or has serious difficulty hearing
R02_AX0191_NB	Adult	Is blind or has serious difficulty seeing, even when wearing glasses
R02_AX0297_OS	Adult	Brand of your favorite e-cigarette advertisement - Something else - Specify
R02_AX0197_OS	Adult	Brand of favorite tobacco advertisement - Specify
R02_AM0012	Adult	Ever been enrolled in VA Health Care
R02_AM0021	Adult	Sexual attraction to gender
R02_AM0061	Adult	Consider yourself to be transgender

Table B-3. Wave 2 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R02_AM0062	Adult	Consider self to be male-to-female, female-to-male, or non-conforming
R02_LCAD01	Adult	Language in which CAPI portions of adult interview were conducted
R02_AX0217_1A	Adult	Tobacco Product 1, bar code scan
R02_AX0217_2A	Adult	Tobacco Product 2, bar code scan
R02_AX0217_3A	Adult	Tobacco Product 3, bar code scan
R02_AX0217_4A	Adult	Tobacco Product 4, bar code scan
R02_AX0217_5A	Adult	Tobacco Product 5, bar code scan
R02_AX0217_6A	Adult	Tobacco Product 6, bar code scan
R02_AX0217_7A	Adult	Tobacco Product 7, bar code scan
R02_AX0217_8A	Adult	Tobacco Product 8, bar code scan
R02_AX0217_9A	Adult	Tobacco Product 9, bar code scan
R02_AX0217_10A	Adult	Tobacco Product 10, bar code scan
R02_CPT05	Parent	Confirm DOB for Wave 1 shadow youth who is expected to be a youth at Wave 2
R02_CPT05C	Parent	Confirm DOB for Wave 1 youth who is expected to be a youth at Wave 2
R02_CPT05D	Parent	Confirm DOB for Wave 1 shadow youth who is expected to be a shadow youth at Wave 2
R02_CPT07	Parent	Youth date of birth (Corrected)
R02_PT0046_NEW_FNAME	Parent	First name of spouse/partner that lives with parent in DU - Specify
R02_PT0046_NEW_LNAME	Parent	Last name of spouse/partner that lives with parent in DU - Specify
R02_PM0058_NEW_FNAME	Parent	First name of other parental figure or guardian - Specify
R02_PM0058_NEW_LNAME	Parent	Last name of other parental figure or guardian - Specify
R02_PM0021_NB	Parent	First name of sibling that youth is a twin of
R02_PM0030_NB	Parent	First names of siblings in multiple birth that are identical to youth
R02_PM0035_NB	Parent	First names of siblings that are in the multiple birth with youth
R02_PARENT_PERSONID	Parent	Parent/guardian Participant ID Number
R02_PM0053	Parent	Confirm parent's relationship to youth
R02_PT0001_OS_1	Parent	Parent or guardian relationship to youth: Other relative
R02_PT0001_OS_2	Parent	Parent or guardian relationship to youth: Other non-relative
R02_PM0052	Parent	Confirm parent's spouse is same as reported in Wave 1
R02_PT0046_PERSONID	Parent	Parent's spouse/partner's participant ID number (if different from Wave 1)
R02_PT0046_NEW_AGE	Parent	Age of new spouse/partner
R02_PT0002_OS_1	Parent	Parent's spouse or partner relationship to youth: Other relative - specify
R02_PT0002_OS_2	Parent	Parent's spouse or partner relationship to youth: Non-relative - specify
R02_PM0058_PERSONID	Parent	Other parental figure/guardian participant ID number
R02_PM0058_NEW_AGE	Parent	Age of other parental figure/guardian
R02_PM0059_OS_1	Parent	Other parental figure/guardian's relationship to sampled youth: Other relative - specify

Table B-3. Wave 2 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R02_PM0059_OS_2	Parent	Other parental figure/guardian's relationship to sampled youth: Other non-relative - specify
R02_PM0016_NB	Parent	Youth was identified as a twin or part of a multiple birth
R02_PM0017_NB	Parent	Youth and sibling are identical twins
R02_PM0019_NB	Parent	Any sibling in multiple birth identical to youth
R02_PX0186_NB	Parent	Youth has serious difficulty walking or climbing stairs
R02_PX0191_NB	Parent	Youth is blind or has serious difficulty seeing even when wearing glasses
R02_PX0190_NB	Parent	Youth is deaf or has serious difficulty hearing
R02_LCPT01	Parent	Language in which the CAPI portions of the interview was completed (interviewer report)
R02_E_YOUTH_LD	Youth	Emancipated youth
R02_CONTINUING_EYOUTH_LD	Youth	Emancipated Youth respondent at Wave 1 is an Emancipated Youth respondent at Wave 2
R02_NEW_BASELINE_EYOUTH_LD	Youth	Youth respondent at Wave 1 is an Emancipated Youth respondent at Wave 2
R02_LYH01	Youth	Youth's preferred language to complete ACASI interview
R02_LEY01	Youth	Emancipated Youth's preferred language to complete ACASI interview
R02_YC1118_OS	Youth	In past 30 days, how you usually got your own cigarettes - Specify
R02_YC1033_OS	Youth	Retail location where your cigarettes are bought most of the time - Specify
R02_YC1049	Youth	Brand of cigarettes usually/last smoked - Specify
R02_YC1071	Youth	Sub-brand of cigarette product usually/last smoked - Specify
R02_YO9035_OS	Youth	Ever used some other electronic nicotine product - Specify
R02_YE1118_OS	Youth	In past 30 days, how you usually got your own e-cigarettes/cartridges and e-liquid - Specify
R02_YE1033_OS	Youth	Retail location where your e-cigarettes/cartridges and e-liquid are bought most of the time - Specify
R02_YE1049	Youth	Brand of e-cigarettes/cartridges/e-liquid usually/last used - Specify
R02_YG1118TC_OS	Youth	In past 30 days, how you usually got your own traditional cigars - Specify
R02_YG1033TC_OS	Youth	Retail location where your traditional cigars are bought most of the time - Specify
R02_YG1049TC	Youth	Brand of traditional cigars usually/last smoked - Specify
R02_YG1071TC	Youth	Sub-brand of traditional cigar product usually/last smoked - Specify
R02_YG1118CL_OS	Youth	In past 30 days, how you usually got your own cigarillos - Specify
R02_YG1033CL_OS	Youth	Retail location where your cigarillos are bought most of the time - Specify
R02_YG1049CL	Youth	Brand of cigarillos usually/last smoked - Specify
R02_YG1071CL	Youth	Sub-brand of cigarillo product usually/last smoked - Specify
R02_YG1118FC_OS	Youth	In past 30 days, how you usually got your own filtered cigars - Specify

Table B-3. Wave 2 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R02_YG1033FC_OS	Youth	Retail location where your filtered cigars are bought most of the time - Specify
R02_YG1049FC	Youth	Brand of filtered cigars usually/last smoked - Specify
R02_YG1071FC	Youth	Sub-brand of filtered cigar product usually/last smoked - Specify
R02_YG1118TJ_OS	Youth	In past 30 days, how you usually got your own traditional cigars as a blunt - Specify
R02_YG1033TJ_OS	Youth	Retail location where your traditional cigars for blunts are bought most of the time - Specify
R02_YG1049TJ	Youth	Brand of traditional cigars usually/last smoked as a blunt - Specify
R02_YG1071TJ	Youth	Sub-brand of traditional cigar product usually/last smoked as a blunt - Specify
R02_YG1118GJ_OS	Youth	In past 30 days, how you usually got your own cigarillo as a blunt - Specify
R02_YG1033GJ_OS	Youth	Retail location where your cigarillos for blunts are bought most of the time - Specify
R02_YG1049GJ	Youth	Brand of cigarillos usually/last smoked as a blunt - Specify
R02_YG1071GJ	Youth	Sub-brand of cigarillo product usually/last smoked as a blunt - Specify
R02_YG1118FJ_OS	Youth	In past 30 days, how you usually got your own filtered cigars as a blunt - Specify
R02_YG1033FJ_OS	Youth	Retail location where your filtered cigars for blunts are bought most of the time - Specify
R02_YG1049FJ	Youth	Brand of filtered cigars usually/last smoked as a blunt - Specify
R02_YG1071FJ	Youth	Sub-brand of filtered cigar product usually/last smoked as a blunt - Specify
R02_YP1118_OS	Youth	In past 30 days, how you usually got your own pipe tobacco - Specify
R02_YP1033_OS	Youth	Retail location where your pipe tobacco are bought most of the time - Specify
R02_YH9011_OS	Youth	Usually smoke hookah: Somewhere else - Specify
R02_YH1118_OS	Youth	In past 30 days, how you usually got your own shisha or hookah tobacco - Specify
R02_YH1033_OS	Youth	Retail location where your shisha or hookah tobacco is bought most of the time - Specify
R02_YH1049	Youth	Brand of hookah tobacco usually/last smoked - Specify
R02_YH1071	Youth	Sub-brand of hookah tobacco product usually/last smoked - Specify
R02_YH1073	Youth	Smoke hookah because: It is part of my cultural tradition
R02_YS1118SU_OS	Youth	In past 30 days, how you usually got your own snus pouches - Specify
R02_YS1033SU_OS	Youth	Retail location where your snus pouches are bought most of the time - Specify
R02_YS1049SU	Youth	Brand of snus pouches usually/last used - Specify
R02_YS1071SU	Youth	Sub-brand of snus pouch product usually/last used - Specify

Table B-3. Wave 2 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R02_YS1118SM_OS	Youth	In past 30 days, how you usually got your own smokeless tobacco - Specify
R02_YS1033SM_OS	Youth	Retail location where your smokeless tobacco is bought most of the time - Specify
R02_YS1049SM	Youth	Brand of smokeless tobacco usually/last used - Specify
R02_YS1071SM	Youth	Sub-brand of smokeless tobacco product usually/last used - Specify
R02_YY0601_OS	Youth	First type of tobacco you tried - Specify
R02_YX0302	Youth	In past 12 months, visited an emergency room or urgent care center for a health problem, accident or injury
R02_YT0034	Youth	In past 12 months, number of visits to emergency room or urgent care center
R02_YX0186_NB	Youth	Has serious difficulty walking or climbing stairs
R02_YX0188_NB	Youth	Has difficulty dressing or bathing
R02_YX0191_NB	Youth	Is blind or has serious difficulty seeing, even when wearing glasses
R02_YX0190_NB	Youth	Is deaf or has serious difficulty hearing
R02_YX0189_NB	Youth	Has serious difficulty concentrating, remembering or making decisions because of a physical, mental or emotional condition
R02_YX0187_NB	Youth	Has difficulty doing errands alone such as visiting a doctor's office or shopping because of a physical, mental or emotional condition
R02_YT0050_12M	Youth	In past 12 months, been told by a doctor, nurse or other health professional that you have high blood pressure
R02_YT0050_NB	Youth	Ever been told by a doctor, nurse or other health professional that you have high blood pressure
R02_YT0041_NB	Youth	Age when you were first told you had high blood pressure
R02_YT0253_NB	Youth	Nonresponse follow-up probe: Age group you were in when you were first told you had high blood pressure
R02_YT0051_12M	Youth	In past 12 months, been told by a doctor, nurse or other health professional that you have high cholesterol
R02_YT0051_NB	Youth	Ever been told by a doctor, nurse or other health professional that you have high cholesterol
R02_YT0043_NB	Youth	Age when you were first told you had high cholesterol
R02_YT0254_NB	Youth	Nonresponse follow-up probe: Age group you were in when you were first told you had high cholesterol
R02_YT0031_12M	Youth	In past 12 months, been told by a doctor, nurse or other health professional that you have asthma
R02_YT0033_12M	Youth	In past 12 months, been told by a doctor, nurse or other health professional that you have bronchitis, pneumonia or chronic cough
R02_YT0040_NB	Youth	Ever been told by a doctor, nurse or other health professional that you have bronchitis, pneumonia or chronic cough
R02_YT0052_NB	Youth	Ever been told by a doctor, nurse or other health professional that you have ADHD or ADD
R02_YT0052_12M	Youth	In past 12 months, been told by a doctor, nurse or other health professional that you have ADHD or ADD

Table B-3. Wave 2 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R02_YT0048	Youth	In past 12 months, took medications regularly for ADHD or ADD
R02_YT0044_NB	Youth	Ever been told by a doctor, dentist or other health professional that you have dental health issues
R02_YT0049_12M	Youth	In past 12 months, been told by a doctor, dentist or other health professional that you have dental health issues
R02_YT0281_NB	Youth	Ever been told by a doctor, nurse or other health professional that you have diabetes, sugar diabetes, high blood sugar or borderline diabetes
R02_YT0042_NB	Youth	Age when you were first told you had diabetes, sugar diabetes, high blood sugar or borderline diabetes
R02_YT0263_NB	Youth	Nonresponse follow-up probe: Age group you were in when you were first told you had diabetes, sugar diabetes, high blood sugar or borderline diabetes
R02_YT0281_12M	Youth	In past 12 months, been told by a doctor, nurse or other health professional that you have diabetes, sugar diabetes, high blood sugar or borderline diabetes
R02_YX0136	Youth	Currently pregnant
R02_YX0137_NN	Youth	Number of weeks/months pregnant - Number
R02_YX0137_UN	Youth	Number of weeks/months pregnant - Unit
R02_YX0478_OS	Youth	Received discount coupons from: Some other way - Specify
R02_YX0480_OS	Youth	How tobacco company sent you the information other than coupons: Some other way - Specify
R02_YX0297_OS	Youth	Brand of your favorite e-cigarette advertisement - Something else - Specify
R02_YX0197_OS	Youth	Brand of favorite tobacco advertisement - Specify
R02_YM0020	Youth	Last grade/year in school completed
R02_YT0019	Youth	In past 12 months, academic performance at school
R02_YT0030	Youth	In past 12 months, how often missed school due to illness
R02_YM0030	Youth	In past 12 months, total household income category
R02_YM0031	Youth	In past 12 months, total household income above or below $50,000
R02_YL0040	Youth	Home is owned or rented
R02_YM0021	Youth	Sexual attraction to gender
R02_YM0063	Youth	Sexual orientation
R02_YM0061	Youth	Transgender
R02_YM0062	Youth	Transgender category
R02_LCYS01	Youth	Language in which CAPI portions of youth interview were conducted

Table B-4. Wave 2 questionnaire variables excluded from the PUFs that are replaced with a derived variable

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R02_AM0001 R02_AM0002 R02_AM0003	Adult	R02R_A_AGECAT7	Age range when interviewed (7 levels)
R02_AC1007_NB R02_AC1121_NB	Adult	R02R_A_AC1007_NB	Age range when first started smoking cigarettes fairly regularly (6 levels)
R02_AC1020_NB R02_AC1122_NB	Adult	R02R_A_AC1020_NB	Age range when first started smoking cigarettes every day (6 levels)
R02_AE1007_NB R02_AE1121_NB	Adult	R02R_A_AE1007_NB	Age range when first started using e-cigarettes fairly regularly (6 levels)
R02_AE1020_NB R02_AE1122_NB	Adult	R02R_A_AE1020_NB	Age range when first started using e-cigarettes every day (6 levels)
R02_AG1007TC_NB R02_AG1121TC_NB	Adult	R02R_A_AG1007TC_NB	Age range when first started smoking traditional cigars fairly regularly (6 levels)
R02_AG1020TC_NB R02_AG1122TC_NB	Adult	R02R_A_AG1020TC_NB	Age range when first started smoking cigarillos every day (6 levels)
R02_AG1007CG_NB R02_AG1121CG_NB	Adult	R02R_A_AG1007CG_NB	Age range when first started smoking filtered cigars fairly regularly (6 levels)
R02_AG1020CG_NB R02_AG1122CG_NB	Adult	R02R_A_AG1020CG_NB	Age range when first started smoking cigarillos every day (6 levels)
R02_AG1007FC_NB R02_AG1121FC_NB	Adult	R02R_A_AG1007FC_NB	Age range when first started smoking filtered cigars fairly regularly (6 levels)
R02_AG1020FC_NB R02_AG1122FC_NB	Adult	R02R_A_AG1020FC_NB	Age range when first started smoking filtered cigars every day (6 levels)
R02_AP1007_NB R02_AP1121_NB	Adult	R02R_A_AP1007_NB	Age range when first started smoking a pipe filled with tobacco fairly regularly (6 levels)
R02_AP1020_NB R02_AP1122_NB	Adult	R02R_A_AP1020_NB	Age range when first started smoking a pipe filled with tobacco every day (6 levels)
R02_AH1007_NB R02_AH1121_NB	Adult	R02R_A_AH1007_NB	Age range when first started smoking hookah fairly regularly (6 levels)
R02_AH1020_NB R02_AH1122_NB	Adult	R02R_A_AH1020_NB	Age range when first started smoking hookah every day (6 levels)
R02_AS1007SU_NB R02_AS1121SU_NB	Adult	R02R_A_AS1007SU_NB	Age range when first started using snus pouches fairly regularly (6 levels)
R02_AS1020SU_NB R02_AS1122SU_NB	Adult	R02R_A_AS1020SU_NB	Age range when first started using snus pouches every day (6 levels)
R02_AS1007SM_NB R02_AS1121SM_NB	Adult	R02R_A_AS1007SM_NB	Age range when first started using smokeless tobacco fairly regularly (6 levels)
R02_AS1020SM_NB R02_AS1122SM_NB	Adult	R02R_A_AS1020SM_NB	Age range when first started using smokeless tobacco every day (6 levels)

Table B-4. Wave 2 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R02_AD1007_NB R02_AD1121_NB	Adult	R02R_A_AD1007_NB	Age range when first started using dissolvable tobacco fairly regularly (6 levels)
R02_AD1020_NB R02_AD1122_NB	Adult	R02R_A_AD1020_NB	Age range when first started using dissolvable tobacco every day (6 levels)
R02_AX0114_NB R02_AX0253_NB	Adult	R02R_A_AX0114_NB	Age range when you were first told you had high blood pressure (6 levels)
R02_AX0115_NB R02_AX0254_NB	Adult	R02R_A_AX0115_NB	Age range when you were first told you had high cholesterol (6 levels)
R02_AX0116_NB R02_AX0255_NB	Adult	R02R_A_AX0116_NB	Age range when you were first told you had congestive heart failure (6 levels)
R02_AX0117_NB R02_AX0256_NB	Adult	R02R_A_AX0117_NB	Age range when you were first told you had a stroke (6 levels)
R02_AX0112_NB R02_AX0252_NB	Adult	R02R_A_AX0112_NB	Age range when you were first told you had a heart attack (6 levels)
R02_AX0120_NB R02_AX0257_NB	Adult	R02R_A_AX0120_NB	Age range when you were first told you had COPD (6 levels)
R02_AX0121_NB R02_AX0258_NB	Adult	R02R_A_AX0121_NB	Age range when you were first told you had chronic bronchitis (6 levels)
R02_AX0123_NB R02_AX0259_NB	Adult	R02R_A_AX0123_NB	Age range when you were first told you had emphysema (6 levels)
R02_AX0124_NB R02_AX0260_NB	Adult	R02R_A_AX0124_NB	Age range when you were first told you had asthma (6 levels)
R02_AX0131_NB R02_AX0261_NB	Adult	R02R_A_AX0131_NB	Age range when you were first told you had gum disease (6 levels)
R02_AX0133_NB R02_AX0262_NB	Adult	R02R_A_AX0133_NB	Age range when you were first told you had pre-cancerous oral lesions (6 levels)
R02_AX0280_NB R02_AX0263_NB	Adult	R02R_A_AX0280_NB	Age range when you were first told you had diabetes, sugar diabetes, high blood sugar or borderline diabetes (6 levels)
R02_AX0143_NB R02_AX0264_NB	Adult	R02R_A_AX0143_NB	Age range when you were first told you had an ulcer (6 levels)
R02_AX0148_NB R02_AX0266_NB	Adult	R02R_A_AX0148_NB	Age range when you were first told you had stomach or gastro-intestinal bleeding (6 levels)
R02_AX0150_NB R02_AX0267_NB	Adult	R02R_A_AX0150_NB	Age range when you were first told you had osteoporosis (6 levels)
R02_AX0198_NB R02_AX0268_NB	Adult	R02R_A_AX0198_NB	Age range when you were first told you had a bone fracture because you have fragile bones (6 levels)
R02_AX0152_NB R02_AX0269_NB	Adult	R02R_A_AX0152_NB	Age range when you were first told you had a cataract or glaucoma (6 levels)
R02_AX0145_02 R02_AX0145_03 R02_AX0145_04 R02_AX0145_05 R02_AX0145_09	Adult	R02R_A_AX0145_NONTOB	Type of cancer is a non-tobacco related cancer (2 levels)

Table B-4. Wave 2 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R02_AX0145_12 R02_AX0145_15 R02_AX0145_18 R02_AX0145_19 R02_AX0145_21 R02_AX0145_16 R02_AX0145_23 R02_AX0145_24 R02_AX0145_25 R02_AX0145_27 R02_AX0145_29 R02_AX0145_30 R02_AX0145_31			
R02_AX0145_01 R02_AX0145_06 R02_AX0145_07 R02_AX0145_08 R02_AX0145_10 R02_AX0145_11 R02_AX0145_13 R02_AX0145_14 R02_AX0145_17 R02_AX0145_20 R02_AX0145_22 R02_AX0145_26 R02_AX0145_28	Adult	R02R_A_AX0145_TOB	Type of cancer is a tobacco related cancer (2 levels)
R02_AX0146_NB_02 R02_AX0265_NB_02 R02_AX0146_NB_03 R02_AX0265_NB_03 R02_AX0146_NB_04 R02_AX0265_NB_04 R02_AX0146_NB_05 R02_AX0265_NB_05 R02_AX0146_NB_09 R02_AX0265_NB_09 R02_AX0146_NB_12 R02_AX0265_NB_12 R02_AX0146_NB_15 R02_AX0265_NB_15 R02_AX0146_NB_18 R02_AX0265_NB_18 R02_AX0146_NB_19 R02_AX0265_NB_19 R02_AX0146_NB_21	Adult	R02R_A_AX0146_NONTOB	Age range when nontobacco related cancer was diagnosed (6 levels)

Table B-4. Wave 2 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R02_AX0265_NB_21 R02_AX0146_NB_16 R02_AX0265_NB_16 R02_AX0146_NB_23 R02_AX0265_NB_23 R02_AX0146_NB_24 R02_AX0265_NB_24 R02_AX0146_NB_25 R02_AX0265_NB_25 R02_AX0146_NB_27 R02_AX0265_NB_27 R02_AX0146_NB_29 R02_AX0265_NB_29 R02_AX0146_NB_30 R02_AX0265_NB_30 R02_AX0146_NB_31 R02_AX0265_NB_31			
R02_AX0146_NB_01 R02_AX0265_NB_01 R02_AX0146_NB_06 R02_AX0265_NB_06 R02_AX0146_NB_07 R02_AX0265_NB_07 R02_AX0146_NB_08 R02_AX0265_NB_08 R02_AX0146_NB_10 R02_AX0265_NB_10 R02_AX0146_NB_11 R02_AX0265_NB_11 R02_AX0146_NB_13 R02_AX0265_NB_13 R02_AX0146_NB_14 R02_AX0265_NB_14 R02_AX0146_NB_17 R02_AX0265_NB_17 R02_AX0146_NB_20 R02_AX0265_NB_20 R02_AX0146_NB_22 R02_AX0265_NB_22 R02_AX0146_NB_26 R02_AX0265_NB_26 R02_AX0146_NB_28 R02_AX0265_NB_28	Adult	R02R_A_AX0146_TOB	Age range when tobacco related cancer was diagnosed (6 levels)
R02_AX0135_12M	Adult	R02R_A_AX0135_12M	Outcome of last pregnancy was miscarriage, induced abortion, ectopic or tubal pregnancy or stillbirth (2 levels)

Table B-4. Wave 2 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R02_AX0300_12M_01 R02_AX0300_12M_02 R02_AX0300_12M_03 R02_AX0300_12M_04 R02_AX0300_12M_05 R02_AX0300_12M_06 R02_AX0300_12M_07	Adult	R02R_A_AX0300_12M	Outcome for live birth was preterm birth, low birth weight, birth defects, placenta previa, placenta abruption, or preeclampsia (2 levels)
R02_AM0018	Adult	R02R_A_AM0018	Highest grade or level of school completed (6 levels)
R02_AM0030	Adult	R02R_A_AM0030	Total household income in the past 12 months (5 levels)
R02_AM0038	Adult	R02R_A_AM0038	Type of current residence (7 levels)
R02_AM0033	Adult	R02R_A_AM0033	Highest grade or year of school completed by mother, step-mother or mother-figure (6 levels)
R02_AM0034	Adult	R02R_A_AM0034	Highest grade or year of school completed by father, step-father or father-figure (6 levels)
R02_AM0036	Adult	R02R_A_AM0036	In past 12 months, parents' total household income (5 levels)
R02_AM0026_01 R02_AM0026_02 R02_AM0026_03 R02_AM0026_04 R02_AM0026_05 R02_AM0026_06 R02_AM0026_07 R02_AM0026_08	Adult	R02R_A_AM0026	Currently covered by health insurance or health coverage plan (5 levels)
R02_AM0063	Adult	R02R_A_SEXORIENT2	Adult sexual orientation (2 levels)
R02_AX0313 R02_AX0679_FT R02_AX0679_IN R02_AX0316 R02_AX0314 R02_AX0109 R02_AX0312	Adult	R02R_A_BMI	Body mass index
R02_PT0001	Parent	R02R_Y_PT0001	Parent or guardian relationship to youth (4 levels)
R02_PT0047	Parent	R02R_Y_PT0047	Parent or guardian marital status (3 levels)
R02_PT0045 R02_PM0057	Parent	R02R_P_OTHPAR_INHH	Youth has any parental figures/guardians other than the parent/guardian in the house
R02_PT0002	Parent	R02R_Y_PT0002	Parent's spouse or partner relationship to youth (4 levels)
R02_PM0059	Parent	R02R_Y_PM0059	Other parental figure/guardian's relationship to youth (4 levels)

Table B-4. Wave 2 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R02_PT0041_NB R02_PT0253_NB	Parent	R02R_Y_PT0041_NB	Age range youth was first told he/she had high blood pressure (3 levels)
R02_PT0043_NB R02_PT0254_NB	Parent	R02R_Y_PT0043_NB	Age range youth was first told he/she has high cholesterol (3 levels)
R02_PT0038_NB R02_PT0260_NB	Parent	R02R_Y_PT0038_NB	Age range youth was first told he/she has asthma (3 levels)
R02_PT0042_NB R02_PT0263_NB	Parent	R02R_Y_PT0042_NB	Age range youth was first told by a doctor or other health professional that he/she has diabetes, sugar diabetes, high blood sugar or borderline diabetes (3 levels)
R02_PM0001	Parent	R02R_Y_PM0001	Highest grade or year of school completed by parent (6 levels)
R02_PM0118	Parent	R02R_Y_PM0118	Highest grade or year of school completed by spouse/guardian (6 levels)
R02_PM0130	Parent	R02R_Y_PM0130	Total household income in past 12 months (5 levels)
R02_YM0004_NB	Youth	R02R_Y_SEX	Youth respondent sex
R02_YM0005_NB_01 R02_YM0005_NB_02 R02_YM0005_NB_03 R02_YM0005_NB_04 R02_YM0005_NB_05	Youth	R02R_Y_HISP	Hispanic origin from the interview (2 levels)
R02_YM0006_NB_01 R02_YM0006_NB_02 R02_YM0006_NB_03 R02_YM0006_NB_04 R02_YM0006_NB_05 R02_YM0006_NB_06 R02_YM0006_NB_07 R02_YM0006_NB_08 R02_YM0006_NB_09 R02_YM0006_NB_10 R02_YM0006_NB_11 R02_YM0006_NB_12 R02_YM0006_NB_13 R02_YM0006_NB_14	Youth	R02R_Y_RACECAT3	Race from the interview (3 levels)
R02_YC1006_NB R02_YC1120_NB	Youth	R02R_Y_YC1006_NB	Age range when first tried cigarette smoking, even one or two puffs (3 levels)
R02_YC1007 R02_YC1121	Youth	R02R_Y_YC1007	Age range when first started smoking cigarettes fairly regularly (3 levels)
R02_YE1006_NB R02_YE1120_NB	Youth	R02R_Y_YE1006_NB	Age range when first tried an e-cigarette, even one or two times (3 levels)
R02_YE1007 R02_YE1121	Youth	R02R_Y_YE1007	Age range when first started using e-cigarettes fairly regularly (3 levels)
R02_YG1006TC_NB R02_YG1120TC_NB	Youth	R02R_Y_YG1006TC_NB	Age range when first tried a traditional cigar, even one or two puffs (3 levels)
R02_YG1007TC R02_YG1121TC	Youth	R02R_Y_YG1007TC	Age range when first started smoking traditional cigars fairly regularly (3 levels)

Table B-4. Wave 2 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R02_YG1006CL_NB R02_YG1120CL_NB	Youth	R02R_Y_YG1006CL_NB	Age range when first tried a cigarillo, even one or two puffs (3 levels)
R02_YG1007CL R02_YG1121CL	Youth	R02R_Y_YG1007CL	Age range when first started smoking cigarillos fairly regularly (3 levels)
R02_YG1006FC_NB R02_YG1120FC_NB	Youth	R02R_Y_YG1006FC_NB	Age range when first tried a filtered cigar, even one or two puffs (3 levels)
R02_YG1007FC R02_YG1121FC	Youth	R02R_Y_YG1007FC	Age range when first started smoking filtered cigars fairly regularly (3 levels)
R02_YH1006_NB R02_YH1120_NB	Youth	R02R_Y_YH1006_NB	Age range when first tried smoking hookah, even one or two puffs (3 levels)
R02_YH1007 R02_YH1121	Youth	R02R_Y_YH1007	Age range when first started smoking a hookah fairly regularly (3 levels)
R02_YS1006SU_NB R02_YS1120SU_NB	Youth	R02R_Y_YS1006SU_NB	Age range when first tried snus pouches, even one or two times (3 levels)
R02_YS1007SU R02_YS1121SU	Youth	R02R_Y_YS1007SU	Age range when first started using snus pouches fairly regularly (3 levels)
R02_YS1006SM_NB R02_YS1120SM_NB	Youth	R02R_Y_YS1006SM_NB	Age range when first tried smokeless tobacco, even one or two times (3 levels)
R02_YS1007SM R02_YS1121SM	Youth	R02R_Y_YS1007SM	Age range when first started using smokeless tobacco fairly regularly (3 levels)
R02_YX0671_01 R02_YX0671_02 R02_YX0671_03 R02_YX0671_04 R02_YX0671_05	Youth	R02R_Y_YX0671	Anyone who lives with you now use tobacco (3 levels)
R02_YX0086_NB R02_YX0087_NB	Youth	R02R_Y_YX0086_NB	Age range when first drank alcohol at all, counting small tastes or sips (3 levels)
R02_YT0038_NB R02_YT0260_NB	Youth	R02R_Y_YT0038_NB	Age range when you were first told you had asthma (3 levels)
R02_YX0074_NB R02_YX0270_NB	Youth	R02R_Y_YX0074_NB	Age range when had first alcoholic drink, other than small tastes or sips (3 levels)
R02_YX0079_NB R02_YX0271_NB	Youth	R02R_Y_YX0079_NB	Age range when first used marijuana, hash, THC or grass (3 levels)
R02_YX0082_NB_01 R02_YX0272_NB_01	Youth	R02R_Y_YX0082_NB_01	Age range when first used: Ritalin or Adderall (3 levels)
R02_YX0082_NB_02 R02_YX0272_NB_02	Youth	R02R_Y_YX0082_NB_02	Age range when first used: Painkillers, sedatives or tranquilizers (3 levels)
R02_YX0082_NB_03 R02_YX0272_NB_03	Youth	R02R_Y_YX0082_NB_03	Age range when first used: Cocaine or crack (3 levels)
R02_YX0082_NB_04 R02_YX0272_NB_04	Youth	R02R_Y_YX0082_NB_04	Age range when first used: Stimulants like methamphetamine or speed (3 levels)
R02_YX0272_NB_05 R02_YX0082_NB_05	Youth	R02R_Y_YX0082_NB_05	Age range when first used: Any other drugs like heroin, inhalants, solvents or hallucinogens (3 levels)

Table B-4. Wave 2 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire variable	Instrument	PUF derived variable	PUF derived variable description
R02_YM0019 R02_YM0018	Youth	R02R_Y_YM0018	Grade level (If on holiday or break - grade level entering when returning to school) (7 levels)
R02_PT0007_FT R02_PT0007_IN R02_PT0007_MT R02_PT0008_LB R02_PT0008_KG R02_YX0310 R02_YT0007_FT R02_YT0007_IN R02_YX0311 R02_YT0008 R02_YX0312	Youth/ Parent	R02R_Y_BMI	Body mass index

Table B-5. Wave 3 questionnaire variables entirely excluded from the PUFs

Questionnaire Variable	Instrument	Questionnaire Variable Description
R03_CAD10	Adult	Confirm respondent DOB
R03_LAD01	Adult	Respondent's preferred language to complete ACASI interview
R03_AM0007	Adult	Ever served on active duty in the U.S. Armed Forces, Military Reserves or National Guard
R03_AM0011_01	Adult	Branch served when on active duty: Army
R03_AM0011_02	Adult	Branch served when on active duty: Navy
R03_AM0011_03	Adult	Branch served when on active duty: Air Force
R03_AM0011_04	Adult	Branch served when on active duty: Marine Corps
R03_AM0011_05	Adult	Branch served when on active duty: Coast Guard
R03_AM0072_OS	Adult	Language other than English spoken at home - Specify
R03_AZ1002_OS	Adult	Ever used any other tobacco products - Specify
R03_AC1033MC_OS	Adult	Retail location where your cigarette is purchased most of the time - Specify
R03_AC1049MC	Adult	Brand of cigarettes usually/last smoked - Specify
R03_AC1071MC	Adult	Sub-brand of cigarette product usually/last smoked - Specify
R03_AC1033RY_OS	Adult	Retail location where your roll-your-own cigarette tobacco is purchased most of the time - Specify
R03_AC1049RY	Adult	Brand of roll-your-own cigarette tobacco usually/last smoked - Specify
R03_AC1071RY	Adult	Sub-brand of roll-your-own cigarette tobacco product usually/last smoked - Specify
R03_AV9003	Adult	Brand of [primary electronic nicotine product] owned - Specify
R03_AV1011_OS	Adult	Flavor of [primary electronic nicotine product] when first started using - Specify
R03_AV1131_OS	Adult	In past 30 days, [electronic nicotine products/cartridges/e-liquid] flavor used - Specify
R03_AV1033_OS	Adult	Retail location where your [electronic nicotine products/cartridges/e-liquid] are bought most of the time - Specify
R03_AV1012_OS	Adult	Flavor of regular brand/brand last used - Specify
R03_AV1049	Adult	Brand of [electronic nicotine products/cartridges/e-liquid] usually/last used - Specify
R03_AV2011_OS	Adult	When first started using [secondary electronic nicotine product]s, flavor used - Specify
R03_AJ1033TC_OS	Adult	Retail location where your traditional cigars for blunts is purchased most of the time - Specify
R03_AJ1049TC	Adult	Brand of traditional cigars for blunts usually/last smoked - Specify
R03_AJ1071TC	Adult	Sub-brand of traditional cigar as a blunt product usually/last smoked - Specify
R03_AJ1033CG_OS	Adult	Retail location where your cigarillos for blunts is purchased most of the time - Specify
R03_AJ1049CG	Adult	Brand of cigarillos for blunts usually/last smoked - Specify

Table B-5. Wave 3 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire Variable	Instrument	Questionnaire Variable Description
R03_AJ1071CG	Adult	Sub-brand of cigarillo as a blunt product usually/last smoked - Specify
R03_AJ1033FC_OS	Adult	Retail location where your filtered cigars for blunts is purchased most of the time - Specify
R03_AJ1049FC	Adult	Brand of filtered cigars for blunts usually/last smoked - Specify
R03_AJ1071FC	Adult	Sub-brand of filtered cigar as a blunt product usually/last smoked - Specify
R03_AG1033TC_OS	Adult	Retail location where your traditional cigars is/were purchased most of the time - Specify
R03_AG1049TC	Adult	Brand of traditional cigars usually/last smoked - Specify
R03_AG1071TC	Adult	Sub-brand of traditional cigar product usually/last smoked - Specify
R03_AG1033CG_OS	Adult	Retail location where your cigarillos is/were purchased most of the time - Specify
R03_AG1049CG	Adult	Brand of cigarillos usually/last smoked - Specify
R03_AG1071CG	Adult	Sub-brand of cigarillo product usually/last smoked - Specify
R03_AG1033FC_OS	Adult	Retail location where your filtered cigars is/were purchased most of the time - Specify
R03_AG1049FC	Adult	Brand of filtered cigars usually/last smoked - Specify
R03_AG1071FC	Adult	Sub-brand of filtered cigar product usually/last smoked - Specify
R03_AP1033_OS	Adult	Retail location where your pipe tobacco is purchased most of the time - Specify
R03_AP1049	Adult	Brand of pipe tobacco usually/last smoked - Specify
R03_AP1071	Adult	Sub-brand of pipe tobacco product usually/last smoked - Specify
R03_AH9011_OS	Adult	Place where usually smoke/smoked a hookah - Specify
R03_AH1033_OS	Adult	Retail location where your hookah tobacco is purchased most of the time - Specify
R03_AH1049	Adult	Brand of shisha or hookah tobacco usually/last smoked - Specify
R03_AH1071	Adult	Sub-brand of hookah tobacco product usually/last smoked - Specify
R03_AU1033_OS	Adult	Retail location where your snus pouches is purchased most of the time - Specify
R03_AU1049	Adult	Brand of snus pouches usually/last used - Specify
R03_AU1071	Adult	Sub-brand of snus pouch product usually/last used - Specify
R03_AS1033_OS	Adult	Retail location where your smokeless tobacco is purchased most of the time - Specify
R03_AS1049	Adult	Brand of smokeless tobacco usually/last used - Specify
R03_AS1071	Adult	Sub-brand of smokeless tobacco product usually/last used - Specify

Table B-5. Wave 3 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire Variable	Instrument	Questionnaire Variable Description
R03_AX0190_NB	Adult	Is deaf or has serious difficulty hearing
R03_AX0191_NB	Adult	Is blind or has serious difficulty seeing, even when wearing glasses
R03_AV9035_OS	Adult	Electronic nicotine product you have used - Specify
R03_AX0203_OS	Adult	In past 30 days, place where noticed e-cigarettes or other electronic nicotine products being advertised - Specify
R03_AX0677_OS	Adult	In past 30 days, noticed cigarettes or other tobacco products being advertised - Specify
R03_AX0686_OS	Adult	In past 12 months, liked or followed brand on social media sites - Specify
R03_AM0067	Adult	Enrolled in High School
R03_AM0068	Adult	Current grade in school
R03_AM0020_OS	Adult	Type of degree program currently enrolled in - Specify
R03_AM0012	Adult	Ever been enrolled in VA Health Care
R03_AM0021	Adult	Sexual attraction to gender
R03_AM0061	Adult	Consider yourself to be transgender
R03_AM0062	Adult	Consider self to be male-to-female, female-to-male, or non-conforming
R03_LCAD01	Adult	Language in which CAPI portions of adult interview were conducted
R03_AX0217_1A	Adult	Tobacco Product 1, bar code scan
R03_AX0217_2A	Adult	Tobacco Product 2, bar code scan
R03_AX0217_3A	Adult	Tobacco Product 3, bar code scan
R03_AX0217_4A	Adult	Tobacco Product 4, bar code scan
R03_AX0217_5A	Adult	Tobacco Product 5, bar code scan
R03_AX0217_6A	Adult	Tobacco Product 6, bar code scan
R03_AX0217_7A	Adult	Tobacco Product 7, bar code scan
R03_AX0217_8A	Adult	Tobacco Product 8, bar code scan
R03_AX0217_9A	Adult	Tobacco Product 9, bar code scan
R03_AX0217_10A	Adult	Tobacco Product 10, bar code scan
R03_AM0001	Adult	Date of birth (Corrected)
R03_AM0002	Adult	Respondent age
R03_CPT05	Parent	Confirm DOB for Wave 1 shadow youth who is expected to be a youth at Wave 3
R03_CPT05C	Parent	Confirm DOB for Wave 1 youth who is expected to be a youth at Wave 3
R03_CPT05D	Parent	Confirm DOB for Wave 1 shadow youth who is expected to be a shadow youth at Wave 3
R03_CPT07	Parent	Corrected Youth DOB
R03_PM0021_NB	Parent	First names of siblings in multiple birth that are identical to youth
R03_PM0030_NB	Parent	First name of sibling that youth is a twin of
R03_PM0035_NB	Parent	First names of siblings that are in the multiple birth with youth
R03_PM0058_NEW_FNAME	Parent	First name of other parental figure or guardian
R03_PM0058_NEW_LNAME	Parent	Last name of other parental figure or guardian
R03_PT0046_NEW_FNAME	Parent	First name of spouse/partner

Table B-5. Wave 3 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire Variable	Instrument	Questionnaire Variable Description
R03_PT0046_NEW_LNAME	Parent	Last name of spouse/partner
R03_PARENT_PERSONID	Parent	Wave 3 Parent/guardian Participant ID Number
R03_PM0053	Parent	Confirm parent's relationship to youth
R03_PT0001_OS_1	Parent	Parent or guardian relationship to youth: Other relative
R03_PT0001_OS_2	Parent	Parent or guardian relationship to youth: Other non-relative
R03_PM0052	Parent	Confirm parent's spouse is same as reported in previous wave
R03_PT0046_NEW_AGE	Parent	Age of new spouse/partner
R03_PT0002_OS_1	Parent	Parent's spouse or partner relationship to youth: Other relative - specify
R03_PT0002_OS_2	Parent	Parent's spouse or partner relationship to youth: Non-relative - specify
R03_PM0058_PERSONID	Parent	Other parental figure/guardian participant ID number
R03_PM0058_NEW_AGE	Parent	Age of other parental figure/guardian
R03_PM0059_OS_1	Parent	Other parental figure/guardian's relationship to sampled youth: Other relative - specify
R03_PM0059_OS_2	Parent	Other parental figure/guardian's relationship to sampled youth: Other non-relative - specify
R03_PM0016_NB	Parent	Youth was identified as a twin or part of a multiple birth
R03_PM0017_NB	Parent	Youth and sibling are identical twins
R03_PM0019_NB	Parent	Any sibling in multiple birth identical to youth
R03_PX0186_NB	Parent	Youth has serious difficulty walking or climbing stairs
R03_PX0191_NB	Parent	Youth is blind or has serious difficulty seeing even when wearing glasses
R03_PX0190_NB	Parent	Youth is deaf or has serious difficulty hearing
R03_LCPT01	Parent	Language in which the CAPI portions of the interview was completed (interviewer report)
R03_E_YOUTH_LD	Youth	Respondent is an emancipated youth
R03_LYH01	Youth	Youth's preferred language to complete ACASI interview
R03_LEY01	Youth	Emancipated Youth's preferred language to complete ACASI interview
R03_YM0066	Youth	How well speak English
R03_YM0070	Youth	How well read English
R03_YM0073	Youth	How well write in English
R03_YM0072_OS	Youth	Language other than English spoken at home: Other language - Specify
R03_YC1118_OS	Youth	In past 30 days, how you usually got your own cigarettes - Specify
R03_YC1033_OS	Youth	Retail location where your cigarettes are bought most of the time - Specify
R03_YC1049	Youth	Brand of cigarettes usually/last smoked - Specify
R03_YC1071	Youth	Sub-brand of cigarette product usually/last smoked - Specify

Table B-5. Wave 3 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire Variable	Instrument	Questionnaire Variable Description
R03_YV9035_OS	Youth	Ever used some other electronic nicotine product - specify
R03_YV1011_OS	Youth	Flavor of first [primary electronic nicotine product] used: Specify.
R03_YV1131_OS	Youth	In past 30 days, used [electronic nicotine products/cartridges/e-liquid] flavored to taste like some other flavor - Specify
R03_YV1118_OS	Youth	In past 30 days, how you usually got your own [electronic nicotine products/cartridges/e-liquid] - Specify
R03_YV1033_OS	Youth	Where[electronic nicotine products/cartridges/e-liquid] is purchased - Specify
R03_YV1049	Youth	Brand of e-liquid usually/last used - Specify
R03_YV2011_OS	Youth	Flavor of first [secondary electronic nicotine product] used - Specify
R03_YJ1118TC_OS	Youth	In past 30 days, how you usually got your own traditional cigars as a blunt - Specify
R03_YJ1033TC_OS	Youth	Retail location where your traditional cigars for blunts are bought most of the time - Specify
R03_YJ1049TC	Youth	Brand of traditional cigars usually/last smoked as a blunt - Specify
R03_YJ1071TC	Youth	Sub-brand of traditional cigar product usually/last smoked as a blunt - Specify
R03_YJ1118CG_OS	Youth	In past 30 days, how you usually got your own cigarillo as a blunt - Specify
R03_YJ1033CG_OS	Youth	Retail location where your cigarillos for blunts are bought most of the time - Specify
R03_YJ1049CG	Youth	Brand of cigarillos usually/last smoked as a blunt - Specify
R03_YJ1071CG	Youth	Sub-brand of cigarillo product usually/last smoked as a blunt - Specify
R03_YJ1118FC_OS	Youth	In past 30 days, how you usually got your own filtered cigars as a blunt - Specify
R03_YJ1033FC_OS	Youth	Retail location where your filtered cigars for blunts are bought most of the time - Specify
R03_YJ1049FC	Youth	Brand of filtered cigars usually/last smoked as a blunt - Specify
R03_YJ1071FC	Youth	Sub-brand of filtered cigar product usually/last smoked as a blunt - Specify
R03_YG1118TC_OS	Youth	In past 30 days, how you usually got your own traditional cigars - Specify
R03_YG1033TC_OS	Youth	Retail location where your traditional cigars are purchased most of the time - Specify
R03_YG1049TC	Youth	Brand of traditional cigars usually/last smoked - Specify
R03_YG1071TC	Youth	Sub-brand of traditional cigar product usually/last smoked - Specify
R03_YG1118CL_OS	Youth	In past 30 days, how you usually got your own cigarillos - Specify

Table B-5. Wave 3 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire Variable	Instrument	Questionnaire Variable Description
R03_YG1033CL_OS	Youth	Retail location where your cigarillos are purchased most of the time - Specify
R03_YG1049CL	Youth	Brand of cigarillos usually / last smoked - Specify
R03_YG1071CL	Youth	Sub-brand of cigarillo product usually/last smoked - Specify
R03_YG1118FC_OS	Youth	In past 30 days, how you usually got your own filtered cigars - Specify
R03_YG1033FC_OS	Youth	Retail location where your filtered cigars are purchased most of the time - Specify
R03_YG1049FC	Youth	Brand of filtered cigars usually/last smoked - Specify
R03_YG1071FC	Youth	Sub-brand of filtered cigar product usually/last smoked - Specify
R03_YP1118_OS	Youth	In past 30 days, how you usually got your own pipe tobacco - Specify
R03_YP1033_OS	Youth	Retail location where your pipe tobacco are bought most of the time - Specify
R03_YH9011_OS	Youth	Usually smoke a hookah: Somewhere else - Specify
R03_YH1118_OS	Youth	In past 30 days, how you usually got your own shisha or hookah tobacco - Specify
R03_YH1033_OS	Youth	Retail location where your hookah tobacco is purchased most of the time - Specify
R03_YH1049	Youth	Brand of hookah tobacco usually/last smoked - Specify
R03_YH1071	Youth	Sub-brand of hookah tobacco product usually/last smoked - Specify
R03_YH1073	Youth	Smoke hookah because: It is part of my cultural tradition
R03_YU1118_OS	Youth	In past 30 days, how you usually got your own snus pouches - Specify
R03_YU1033_OS	Youth	Retail location where your snus pouches are purchased most of the time - Specify
R03_YU1049	Youth	Brand of snus pouches usually / last used - Specify
R03_YU1071	Youth	Sub-brand of snus pouches usually / last used - Specify
R03_YS1118_OS	Youth	In past 30 days, how you usually got your own smokeless tobacco - Specify
R03_YS1033_OS	Youth	Retail location where your smokeless tobacco is bought most of the time - Specify
R03_YS1049	Youth	Brand of smokeless tobacco usually/last used - Specify
R03_YS1071	Youth	Sub-brand of smokeless tobacco product usually/last used - Specify
R03_YZ1002_OS	Youth	Ever used any other tobacco products - Specify
R03_YY0601_OS	Youth	First type of tobacco you tried - Specify
R03_YX0686_OS	Youth	In past 12 months, liked or followed brand on social media sites: Other - Specify
R03_YX0136	Youth	Currently pregnant
R03_YX0137_NN	Youth	Number of weeks/months pregnant - Number
R03_YX0137_UN	Youth	Number of weeks/months pregnant - Unit

Table B-5. Wave 3 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire Variable	Instrument	Questionnaire Variable Description
R03_YX0203_OS	Youth	In past 30 days, has noticed e-cigarettes or other electronic nicotine products being advertised - Specify
R03_YX0677_OS	Youth	In past 30 days, noticed cigarettes or other tobacco products being advertised - Specify
R03_YM0020	Youth	Last grade/year in school completed
R03_YM0021	Youth	Sexual attraction to gender
R03_YM0063	Youth	Sexual orientation
R03_YM0061	Youth	Transgender
R03_YM0062	Youth	Transgender category
R03_LCYS01	Youth	Language in which CAPI portions of youth interview were conducted

Table B-6. Wave 3 questionnaire variables excluded from the PUFs that are replaced with a derived variable

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R03_AM0001 R03_AM0002 R03_AM0003	Adult	R03R_A_AGECAT7	Age range when interviewed (7 levels)
R03_AM0069	Adult	R03R_A_AM0069	Respondent is a citizen of the United States (2 levels)
R03_AM0065	Adult	R03R_A_AM0065	Number of years lived in the United States (3 levels)
R03_AC1007_NB R03_AC1121_NB	Adult	R03R_A_AC1007_NB	Age range when first started smoking cigarettes fairly regularly (6 levels)
R03_AC1020_NB R03_AC1122_NB	Adult	R03R_A_AC1020_NB	Age range when first started smoking cigarettes every day (6 levels)
R03_AV1007_NB R03_AV1121_NB	Adult	R03R_A_AV1007_NB	Age range when first started using [EPRODTYPE1]s fairly regularly (6 levels)
R03_AV1020_NB R03_AV1122_NB	Adult	R03R_A_AV1020_NB	Age range when first started using [EPRODTYPE1]s every day (6 levels)
R03_AG1007TC_NB R03_AG1121TC_NB	Adult	R03R_A_AG1007TC_NB	Age range when first started smoking traditional cigars fairly regularly (6 levels)
R03_AG1020TC_NB R03_AG1122TC_NB	Adult	R03R_A_AG1020TC_NB	Age range when first started smoking cigarillos every day (6 levels)
R03_AG1007CG_NB R03_AG1121CG_NB	Adult	R03R_A_AG1007CG_NB	Age range when first started smoking filtered cigars fairly regularly (6 levels)
R03_AG1020CG_NB R03_AG1122CG_NB	Adult	R03R_A_AG1020CG_NB	Age range when first started smoking cigarillos every day (6 levels)
R03_AG1007FC_NB R03_AG1121FC_NB	Adult	R03R_A_AG1007FC_NB	Age range when first started smoking filtered cigars fairly regularly (6 levels)
R03_AG1020FC_NB R03_AG1122FC_NB	Adult	R03R_A_AG1020FC_NB	Age range when first started smoking filtered cigars every day (6 levels)
R03_AP1007_NB R03_AP1121_NB	Adult	R03R_A_AP1007_NB	Age range when first started smoking a pipe filled with tobacco fairly regularly (6 levels)
R03_AP1020_NB R03_AP1122_NB	Adult	R03R_A_AP1020_NB	Age range when first started smoking a pipe filled with tobacco every day (6 levels)
R03_AH1007_NB R03_AH1121_NB	Adult	R03R_A_AH1007_NB	Age range when first started smoking hookah fairly regularly (6 levels)
R03_AH1020_NB R03_AH1122_NB	Adult	R03R_A_AH1020_NB	Age range when first started smoking hookah every day (6 levels)

Table B-6. Wave 3 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R03_AU1007_NB R03_AU1121_NB	Adult	R03R_A_AU1007_NB	Age range when first started smoking snus pouches fairly regularly (6 levels)
R03_AU1020_NB R03_AU1122_NB	Adult	R03R_A_AU1020_NB	Age range when first started smoking snus pouches every day (6 levels)
R03_AS1007_NB R03_AS1121_NB	Adult	R03R_A_AS1007_NB	Age range when first started using smokeless tobacco fairly regularly (6 levels)
R03_AS1020_NB R03_AS1122_NB	Adult	R03R_A_AS1020_NB	Age range when first started using smokeless tobacco every day (6 levels)
R03_AX0114_NB R03_AX0253_NB	Adult	R03R_A_AX0114_NB	Age range when you were first told you had high blood pressure (6 levels)
R03_AX0115_NB R03_AX0254_NB	Adult	R03R_A_AX0115_NB	Age range when you were first told you had high cholesterol (6 levels)
R03_AX0116_NB R03_AX0255_NB	Adult	R03R_A_AX0116_NB	Age range when you were first told you had congestive heart failure (6 levels)
R03_AX0117_NB R03_AX0256_NB	Adult	R03R_A_AX0117_NB	Age range when you were first told you had a stroke (6 levels)
R03_AX0112_NB R03_AX0252_NB	Adult	R03R_A_AX0112_NB	Age range when you were first told you had a heart attack (6 levels)
R03_AX0120_NB R03_AX0257_NB	Adult	R03R_A_AX0120_NB	Age range when you were first told you had COPD (6 levels)
R03_AX0121_NB R03_AX0258_NB	Adult	R03R_A_AX0121_NB	Age range when you were first told you had chronic bronchitis (6 levels)
R03_AX0123_NB R03_AX0259_NB	Adult	R03R_A_AX0123_NB	Age range when you were first told you had emphysema (6 levels)
R03_AX0124_NB R03_AX0260_NB	Adult	R03R_A_AX0124_NB	Age range when you were first told you had asthma (6 levels)
R03_AX0131_NB R03_AX0261_NB	Adult	R03R_A_AX0131_NB	Age range when you were first told you had gum disease (6 levels)
R03_AX0133_NB R03_AX0262_NB	Adult	R03R_A_AX0133_NB	Age range when you were first told you had pre-cancerous oral lesions (6 levels)
R03_AX0280_NB R03_AX0263_NB	Adult	R03R_A_AX0280_NB	Age range when you were first told you had diabetes, sugar diabetes, high blood sugar or borderline diabetes (6 levels)
R03_AX0143_NB R03_AX0264_NB	Adult	R03R_A_AX0143_NB	Age range when you were first told you had an ulcer (6 levels)
R03_AX0148_NB R03_AX0266_NB	Adult	R03R_A_AX0148_NB	Age range when you were first told you had stomach or gastro-intestinal bleeding (6 levels)
R03_AX0150_NB R03_AX0267_NB	Adult	R03R_A_AX0150_NB	Age range when you were first told you had osteoporosis (6 levels)

Table B-6. Wave 3 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R03_AX0198_NB R03_AX0268_NB	Adult	R03R_A_AX0198_NB	Age range when you were first told you had a bone fracture because you have fragile bones (6 levels)
R03_AX0152_NB R03_AX0269_NB	Adult	R03R_A_AX0152_NB	Age range when you were first told you had a cataract or glaucoma (6 levels)
R03_AX0703 R03_AX0704	Adult	R03R_A_AX0703	Age range when you were first told you had macular degeneration (6 levels)
R03_AX0145_02 R03_AX0145_03 R03_AX0145_04 R03_AX0145_05 R03_AX0145_09 R03_AX0145_12 R03_AX0145_15 R03_AX0145_18 R03_AX0145_19 R03_AX0145_21 R03_AX0145_16 R03_AX0145_23 R03_AX0145_24 R03_AX0145_25 R03_AX0145_27 R03_AX0145_29 R03_AX0145_30 R03_AX0145_31	Adult	R03R_A_AX0145_NONTOB	Type of cancer is a non-tobacco related cancer (2 levels)
R03_AX0145_01 R03_AX0145_06 R03_AX0145_07 R03_AX0145_08 R03_AX0145_10 R03_AX0145_11 R03_AX0145_13 R03_AX0145_14 R03_AX0145_17 R03_AX0145_20 R03_AX0145_22 R03_AX0145_26 R03_AX0145_28	Adult	R03R_A_AX0145_TOB	Type of cancer is a tobacco related cancer (2 levels)

Table B-6. Wave 3 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R03_AX0146_NB_02 R03_AX0265_NB_02 R03_AX0146_NB_03 R03_AX0265_NB_03 R03_AX0146_NB_04 R03_AX0265_NB_04 R03_AX0146_NB_05 R03_AX0265_NB_05 R03_AX0146_NB_09 R03_AX0265_NB_09 R03_AX0146_NB_12 R03_AX0265_NB_12 R03_AX0146_NB_15 R03_AX0265_NB_15 R03_AX0146_NB_18 R03_AX0265_NB_18 R03_AX0146_NB_19 R03_AX0265_NB_19 R03_AX0146_NB_21 R03_AX0265_NB_21 R03_AX0146_NB_16 R03_AX0265_NB_16 R03_AX0146_NB_23 R03_AX0265_NB_23 R03_AX0146_NB_24 R03_AX0265_NB_24 R03_AX0146_NB_25 R03_AX0265_NB_25 R03_AX0146_NB_27 R03_AX0265_NB_27 R03_AX0146_NB_29 R03_AX0265_NB_29 R03_AX0146_NB_30 R03_AX0265_NB_30 R03_AX0146_NB_31 R03_AX0265_NB_31	Adult	R03R_A_AX0146_NONTOB	Age range when nontobacco related cancer was diagnosed (6 levels)

Table B-6. Wave 3 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R03_AX0146_NB_01 R03_AX0265_NB_01 R03_AX0146_NB_06 R03_AX0265_NB_06 R03_AX0146_NB_07 R03_AX0265_NB_07 R03_AX0146_NB_08 R03_AX0265_NB_08 R03_AX0146_NB_10 R03_AX0265_NB_10 R03_AX0146_NB_11 R03_AX0265_NB_11 R03_AX0146_NB_13 R03_AX0265_NB_13 R03_AX0146_NB_14 R03_AX0265_NB_14 R03_AX0146_NB_17 R03_AX0265_NB_17 R03_AX0146_NB_20 R03_AX0265_NB_20 R03_AX0146_NB_22 R03_AX0265_NB_22 R03_AX0146_NB_26 R03_AX0265_NB_26 R03_AX0146_NB_28 R03_AX0265_NB_28	Adult	R03R_A_AX0146_TOB	Age range when tobacco related cancer was diagnosed (6 levels)
R03_AX0135_12M	Adult	R03R_A_AX0135_12M	Outcome of last pregnancy was miscarriage, induced abortion, ectopic or tubal pregnancy or stillbirth (2 levels)
R03_AX0300_12M_01 R03_AX0300_12M_02 R03_AX0300_12M_03 R03_AX0300_12M_04 R03_AX0300_12M_05 R03_AX0300_12M_06 R03_AX0300_12M_07 R03_AX0706	Adult	R03R_A_AX0300_12M	Outcome for live birth was preterm birth, low birth weight, birth defects, placenta previa, placenta abruption, preeclampsia, or cleft lip or palate (2 levels)
R03_AM0018	Adult	R03R_A_AM0018	Highest grade or level of school completed (6 levels)
R03_AM0030	Adult	R03R_A_AM0030	Total household income in the past 12 months (5 levels)
R03_AM0038	Adult	R03R_A_AM0038	Type of current residence (7 levels)
R03_AM0033	Adult	R03R_A_AM0033	Highest grade or year of school completed by mother, step-mother or mother-figure (6 levels)
R03_AM0034	Adult	R03R_A_AM0034	Highest grade or year of school completed by father, step-father or father-figure (6 levels)

Table B-6. Wave 3 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R03_AM0036	Adult	R03R_A_AM0036	In past 12 months, parents' total household income (5 levels)
R03_AM0026_01 R03_AM0026_02 R03_AM0026_03 R03_AM0026_04 R03_AM0026_05 R03_AM0026_06 R03_AM0026_07 R03_AM0026_08	Adult	R03R_A_AM0026	Currently covered by health insurance or health coverage plan (5 levels)
R03_AM0063	Adult	R03R_A_SEXORIENT2	Adult sexual orientation (2 levels)
R03_AX0313 R03_AX0679_FT R03_AX0679_IN R03_AX0316 R03_AX0109 R03_AX0312	Adult	R03R_A_BMI	Body mass index
R03_AT0047	Adult	R03R_A_AT0047	Recoded marital status (3 levels)
R03_PT0001	Parent	R03R_Y_PT0001	Recoded parent or guardian relationship to youth (4 levels)
R03_PT0047	Parent	R03R_Y_PT0047	Recoded parent or guardian marital status (3 levels)
R03_PT0045 R03_PM0057	Parent	R03R_P_OTHPAR_INHH	Recoded youth has any parental figures/guardians other than the parent/guardian in the house (2 levels)
R03_PT0002	Parent	R03R_Y_PT0002	Recoded parent's spouse or partner relationship to youth (4 levels)
R03_PM0059	Parent	R03R_Y_PM0059	Recoded other parental figure/guardian's relationship to youth (4 levels)
R03_PT0041_NB R03_PT0253_NB	Parent	R03R_Y_PT0041_NB	Age range youth was first told he/she had high blood pressure (3 levels)
R03_PT0043_NB R03_PT0254_NB	Parent	R03R_Y_PT0043_NB	Age range youth was first told he/she has high cholesterol (3 levels)
R03_PT0038_NB R03_PT0260_NB	Parent	R03R_Y_PT0038_NB	Age range youth was first told he/she has asthma (3 levels)
R03_PT0042_NB R03_PT0263_NB	Parent	R03R_Y_PT0042_NB	Age range youth was first told by a doctor or other health professional that he/she has diabetes, sugar diabetes, high blood sugar or borderline diabetes (3 levels)
R03_PM0069	Parent	R03R_Y_PM0069	Recoded parent respondent is a citizen of the United States (2 levels)

Table B-6. Wave 3 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R03_PM0065	Parent	R03R_Y_PM0065	Recoded number of years parents lived in the United States (3 levels)
R03_PM0001	Parent	R03R_Y_PM0001	Recoded highest grade or year of school completed by parent (6 levels)
R03_PM0118	Parent	R03R_Y_PM0118	Recoded highest grade or year of school completed by spouse/guardian (6 levels)
R03_PM0130	Parent	R03R_Y_PM0130	Recoded total household income in past 12 months (5 levels)
R03_PL0040	Parent	R03R_Y_PL0040	Recoded home is owned or rented (2 levels)
R03_YM0069	Youth	R03R_Y_YM0069	Recoded citizen of the United States (2 levels)
R03_YM0065	Youth	R03R_Y_YM0065	Recoded number of years youth lived in the United States (3 levels)
R03_YC1006_NB R03_YC1120_NB	Youth	R03R_Y_YC1006_NB	Age range when first tried cigarette smoking, even one or two puffs (3 levels)
R03_YC1007 R03_YC1121	Youth	R03R_Y_YC1007	Age range when first started smoking cigarettes fairly regularly (3 levels)
R03_YV1006_NB R03_YV1120_NB	Youth	R03R_Y_YV1006_NB	Age range when first tried [primary electronic nicotine product], even one or two times, even one or two puffs (3 levels)
R03_YV1007 R03_YV1121	Youth	R03R_Y_YV1007	Age range when first started using [primary electronic nicotine product]s fairly regularly (3 levels)
R03_YG1006TC_NB R03_YG1120TC_NB	Youth	R03R_Y_YG1006TC_NB	Age range when first tried a traditional cigar, even one or two puffs (3 levels)
R03_YG1007TC R03_YG1121TC	Youth	R03R_Y_YG1007TC	Age range when first started smoking traditional cigars fairly regularly (3 levels)
R03_YG1006CL_NB R03_YG1120CL_NB	Youth	R03R_Y_YG1006CL_NB	Age range when first tried a cigarillo, even one or two puffs (3 levels)
R03_YG1007CL R03_YG1121CL	Youth	R03R_Y_YG1007CL	Age range when first started smoking cigarillos fairly regularly (3 levels)
R03_YG1006FC_NB R03_YG1120FC_NB	Youth	R03R_Y_YG1006FC_NB	Age range when first tried a filtered cigar, even one or two puffs (3 levels)
R03_YG1007FC R03_YG1121FC	Youth	R03R_Y_YG1007FC	Age range when first started smoking filtered cigars fairly regularly (3 levels)

Table B-6. Wave 3 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R03_YH1006_NB R03_YH1120_NB	Youth	R03R_Y_YH1006_NB	Age range when first tried smoking hookah, even one or two puffs (3 levels)
R03_YH1007 R03_YH1121	Youth	R03R_Y_YH1007	Age range when first started smoking a hookah fairly regularly (3 levels)
R03_YU1006_NB R03_YU1120_NB	Youth	R03R_Y_YU1006_NB	Age range when first tried snus pouches, even one or two times (3 levels)
R03_YU1007 R03_YU1121	Youth	R03R_Y_YU1007	Age range when first started using snus pouches fairly regularly (3 levels)
R03_YS1006_NB R03_YS1120_NB	Youth	R03R_Y_YS1006_NB	Age range when first tried smokeless tobacco, even one or two times (3 levels)
R03_YS1007 R03_YS1121	Youth	R03R_Y_YS1007	Age range when first started using smokeless tobacco fairly regularly (3 levels)
R03_YX0671_01 R03_YX0671_02 R03_YX0671_03 R03_YX0671_04 R03_YX0671_05 R03_YX0671_06 R03_YX0671_07 R03_YX0671_08	Youth	R03R_Y_YX0671	Recoded anyone who lives with you now use tobacco (4 levels)
R03_YT0038_NB R03_YT0260_NB	Youth	R03R_Y_YT0038_NB	Age range when you were first told you had asthma (3 levels)
R03_YX0086_NB R03_YX0087_NB	Youth	R03R_Y_YX0086_NB	Age range when first drank alcohol at all, counting small tastes or sips (3 levels)
R03_YX0074_NB R03_YX0270_NB	Youth	R03R_Y_YX0074_NB	Age range when had first alcoholic drink, other than small tastes or sips (3 levels)
R03_YX0079_NB R03_YX0271_NB	Youth	R03R_Y_YX0079_NB	Age range when first used marijuana, hash, THC or grass (3 levels)
R03_YX0082_NB_01 R03_YX0272_NB_01	Youth	R03R_Y_YX0082_NB_01	Age range when first used: Ritalin or Adderall (3 levels)
R03_YX0082_NB_02 R03_YX0272_NB_02	Youth	R03R_Y_YX0082_NB_02	Age range when first used: Painkillers, sedatives or tranquilizers (3 levels)
R03_YX0082_NB_03 R03_YX0272_NB_03	Youth	R03R_Y_YX0082_NB_03	Age range when first used: Cocaine or crack (3 levels)
R03_YX0082_NB_04 R03_YX0272_NB_04	Youth	R03R_Y_YX0082_NB_04	Age range when first used: Stimulants like methamphetamine or speed (3 levels)

Table B-6. Wave 3 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R03_YX0082_NB_05 R03_YX0272_NB_05	Youth	R03R_Y_YX0082_NB_05	Age range when first used: Any other drugs like heroin, inhalants, solvents or hallucinogens (3 levels)
R03_YM0019 R03_YM0018	Youth	R03R_Y_YM0018	Recoded grade level (If on holiday or break - grade level entering when returning to school) (7 levels)
R03_YM0004_NB	Youth	R03R_Y_SEX	Youth respondent sex
R03_YM0005_NB_01 R03_YM0005_NB_02 R03_YM0005_NB_03 R03_YM0005_NB_04 R03_YM0005_NB_05	Youth	R03R_Y_HISP	DERIVED - Hispanic origin from the interview (2 levels)
R03_YM0006_NB_01 R03_YM0006_NB_02 R03_YM0006_NB_03 R03_YM0006_NB_04 R03_YM0006_NB_05 R03_YM0006_NB_06 R03_YM0006_NB_07 R03_YM0006_NB_08 R03_YM0006_NB_09 R03_YM0006_NB_10 R03_YM0006_NB_11 R03_YM0006_NB_12 R03_YM0006_NB_13 R03_YM0006_NB_14	Youth	R03R_Y_RACECAT3	Recoded Race from the interview (3 levels)
R03_PT0007_FT R03_PT0007_IN R03_PT0007_MT R03_PT0008_LB R03_PT0008_KG R03_YX0310 R03_YT0007_FT R03_YT0007_IN R03_YX0311 R03_YT0008 R03_YX0312	Youth/ Parent	R03R_Y_BMI	Body mass index

Table B-7. Wave 4 questionnaire variables entirely excluded from the PUFs

Questionnaire variable	Instrument	Questionnaire variable description
R04_CAD10	Adult	Confirm respondent DOB
R04_LAD01	Adult	Respondent's preferred language to complete ACASI interview
R04_AM0007	Adult	Ever served on active duty in the U.S. Armed Forces, Military Reserves or National Guard
R04_AM0011_NB_01	Adult	Branch served when on active duty: Army
R04_AM0011_NB_02	Adult	Branch served when on active duty: Navy
R04_AM0011_NB_03	Adult	Branch served when on active duty: Air Force
R04_AM0011_NB_04	Adult	Branch served when on active duty: Marine Corps
R04_AM0011_NB_05	Adult	Branch served when on active duty: Coast Guard
R04_AM0072_OS	Adult	Language other than English spoken at home - Specify
R04_AZ1002_OS	Adult	Ever used any other tobacco products - Specify
R04_AC1033MC_OS	Adult	Retail location where your cigarette is purchased most of the time - Specify
R04_AC1049MC	Adult	Brand of cigarettes usually/last smoked - Specify
R04_AC1071MC	Adult	Sub-brand of cigarette product usually/last smoked - Specify
R04_AC1033RY_OS	Adult	Retail location where your roll-your-own cigarette tobacco is purchased most of the time - Specify
R04_AC1049RY	Adult	Brand of roll-your-own cigarette tobacco usually/last smoked - Specify
R04_AC1071RY	Adult	Sub-brand of roll-your-own cigarette tobacco product usually/last smoked - Specify
R04_AV9003	Adult	Brand of [primary electronic nicotine product] owned - Specify
R04_AV1011_OS	Adult	Flavor of electronic nicotine product when first started using: Some other flavor - specify
R04_AV9035_OS	Adult	Electronic nicotine product you have used - specify
R04_AV1131_OS	Adult	In past 30 days, [electronic nicotine products/cartridges/e-liquid] flavor used - Specify
R04_AV1033_OS	Adult	Retail location where your [electronic nicotine products/cartridges/e-liquid] are bought most of the time - Specify
R04_AV1012_OS	Adult	Flavor of regular brand/brand last used - Specify
R04_AV1049	Adult	Brand of [electronic nicotine products/cartridges/e-liquid] usually/last used - Specify
R04_AJ1033TC_OS	Adult	Retail location where your traditional cigars for blunts is purchased most of the time - Specify
R04_AJ1049TC	Adult	Brand of traditional cigars for blunts usually/last smoked - Specify
R04_AJ1071TC	Adult	Sub-brand of traditional cigar as a blunt product usually/last smoked - Specify
R04_AJ1033CG_OS	Adult	Retail location where your cigarillos for blunts is purchased most of the time - Specify
R04_AJ1049CG	Adult	Brand of cigarillos for blunts usually/last smoked - Specify
R04_AJ1071CG	Adult	Sub-brand of cigarillo as a blunt product usually/last smoked - Specify

Table B-7. Wave 4 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R04_AJ1033FC_OS	Adult	Retail location where your filtered cigars for blunts is purchased most of the time - Specify
R04_AJ1049FC	Adult	Brand of filtered cigars for blunts usually/last smoked - Specify
R04_AJ1071FC	Adult	Sub-brand of filtered cigar as a blunt product usually/last smoked - Specify
R04_AG1033TC_OS	Adult	Retail location where your traditional cigars is/were purchased most of the time - Specify
R04_AG1049TC	Adult	Brand of traditional cigars usually/last smoked - Specify
R04_AG1071TC	Adult	Sub-brand of traditional cigar product usually/last smoked - Specify
R04_AG1033CG_OS	Adult	Retail location where your cigarillos is/were purchased most of the time - Specify
R04_AG1049CG	Adult	Brand of cigarillos usually/last smoked - Specify
R04_AG1071CG	Adult	Sub-brand of cigarillo product usually/last smoked - Specify
R04_AG1033FC_OS	Adult	Retail location where your filtered cigars is/were purchased most of the time - Specify
R04_AG1049FC	Adult	Brand of filtered cigars usually/last smoked - Specify
R04_AG1071FC	Adult	Sub-brand of filtered cigar product usually/last smoked - Specify
R04_AP1033_OS	Adult	Retail location where your pipe tobacco is purchased most of the time - Specify
R04_AP1049	Adult	Brand of pipe tobacco usually/last smoked - Specify
R04_AP1071	Adult	Sub-brand of pipe tobacco product usually/last smoked - Specify
R04_AH9011_OS	Adult	Place where usually smoke/smoked a hookah - Specify
R04_AH1033_OS	Adult	Retail location where your hookah tobacco is purchased most of the time - Specify
R04_AH1049	Adult	Brand of shisha or hookah tobacco usually/last smoked - Specify
R04_AH1071	Adult	Sub-brand of hookah tobacco product usually/last smoked - Specify
R04_AU1033_OS	Adult	Retail location where your snus pouches is purchased most of the time - Specify
R04_AU1049	Adult	Brand of snus pouches usually/last used - Specify
R04_AU1071	Adult	Sub-brand of snus pouch product usually/last used - Specify
R04_AS1033_OS	Adult	Retail location where your smokeless tobacco is purchased most of the time - Specify
R04_AS1049	Adult	Brand of smokeless tobacco usually/last used - Specify
R04_AS1071	Adult	Sub-brand of smokeless tobacco product usually/last used - Specify
R04_AX0686_OS	Adult	In past 12 months, liked or followed brand on social media sites - Specify
R04_AX0190_NB	Adult	Is deaf or has serious difficulty hearing
R04_AX0191_NB	Adult	Is blind or has serious difficulty seeing, even when wearing glasses
R04_AX0203_OS	Adult	In past 30 days, place where noticed e-cigarettes or other electronic nicotine products being advertised - Specify
R04_AX0677_OS	Adult	In past 30 days, noticed cigarettes or other tobacco products being advertised - Specify

Table B-7. Wave 4 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R04_AM0067	Adult	Enrolled in High School
R04_AM0068	Adult	Current grade in school
R04_AM0020_OS	Adult	Type of degree program currently enrolled in - Specify
R04_AM0012_12M	Adult	Ever been enrolled in VA Health Care
R04_AM0012_NB	Adult	Ever been enrolled in VA Health Care
R04_AM0021	Adult	Sexual attraction to gender
R04_AM0061	Adult	Consider yourself to be transgender
R04_AM0062	Adult	Consider self to be male-to-female, female-to-male, or non-conforming
R04_LCAD01	Adult	Language in which CAPI portions of adult interview were conducted
R04_AX0217_1A	Adult	Tobacco Product 1, bar code scan
R04_AX0217_2A	Adult	Tobacco Product 2, bar code scan
R04_AX0217_3A	Adult	Tobacco Product 3, bar code scan
R04_AX0217_4A	Adult	Tobacco Product 4, bar code scan
R04_AX0217_5A	Adult	Tobacco Product 5, bar code scan
R04_AX0217_6A	Adult	Tobacco Product 6, bar code scan
R04_AX0217_7A	Adult	Tobacco Product 7, bar code scan
R04_AX0217_8A	Adult	Tobacco Product 8, bar code scan
R04_AX0217_9A	Adult	Tobacco Product 9, bar code scan
R04_AX0217_10A	Adult	Tobacco Product 10, bar code scan
R04_AM0001	Adult	Date of birth (Corrected)
R04_AM0002	Adult	Respondent age
R04_AG1011CG_OS	Adult	Flavor of cigarillo when first started smoking: Some other flavor - specify
R04_AG1011FC_OS	Adult	Flavor of filtered cigar when first started smoking: Some other flavor - specify
R04_AG1011TC_OS	Adult	Flavor of traditional cigar when first started smoking: Some other flavor - specify
R04_AG1012CG_OS	Adult	Cigarillo flavor usually/last smoked: Some other flavor - specify
R04_AG1012FC_OS	Adult	Filtered cigar flavor usually/last smoked: Some other flavor - specify
R04_AG1012TC_OS	Adult	Traditional cigar flavor usually/last smoked: Some other flavor - specify
R04_AG1131CG_OS	Adult	In past 30 days, cigarillo flavor smoked: Some other flavor - specify
R04_AG1131FC_OS	Adult	In past 30 days, filtered cigar flavor smoked: Some other flavor - specify
R04_AG1131TC_OS	Adult	In past 30 days, traditional cigar flavor smoked: Some other flavor - specify
R04_AH1011_OS	Adult	Flavor of hookah tobacco when first started smoking: Some other flavor - specify
R04_AH1012_OS	Adult	Hookah tobacco flavor usually/last smoked: Some other flavor - specify
R04_AH1131_OS	Adult	In past 30 days, hookah tobacco flavor smoked: Some other flavor - specify
R04_AJ1011CG_OS	Adult	Flavor of cigarillo used as blunts when first started smoking: Some other flavor - specify
R04_AJ1011FC_OS	Adult	Flavor of filtered cigar used as blunts when first started smoking: Some other flavor - specify

Table B-7. Wave 4 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R04_AJ1011TC_OS	Adult	Flavor of traditional cigar used as blunts when first started smoking: Some other flavor - specify
R04_AJ1012CG_OS	Adult	Cigarillo as blunts flavor usually/last smoked: Some other flavor - specify
R04_AJ1012FC_OS	Adult	Filtered cigar as blunts flavor usually/last smoked: Some other flavor - specify
R04_AJ1012TC_OS	Adult	Traditional cigar as blunts flavor usually/last smoked: Some other flavor - specify
R04_AJ1131CG_OS	Adult	In past 30 days, cigarillo as blunts flavor smoked: Some other flavor - specify
R04_AJ1131FC_OS	Adult	In past 30 days, filtered cigar as blunts flavor smoked: Some other flavor - specify
R04_AJ1131TC_OS	Adult	In past 30 days, traditional cigar as blunts flavor smoked: Some other flavor - specify
R04_AN0336_OS	Adult	Main reason you used a nicotine patch, gum, inhaler, nasal spray, lozenge or pill: Some other reason - specify (current established, current experimental or recent former established non-electronic tobacco users)
R04_AN0336E_OS	Adult	Main reason you used a nicotine patch, gum, inhaler, nasal spray, lozenge or pill: Some other reason - specify (current established, recent former established or current experimental electronic nicotine product users)
R04_AP1011_OS	Adult	Flavor of pipe tobacco when first started smoking: Some other flavor - specify
R04_AP1012_OS	Adult	Pipe tobacco flavor usually/last smoked: Some other flavor - specify
R04_AP1131_OS	Adult	In past 30 days, pipe tobacco flavor smoked: Some other flavor - specify
R04_AS1011_OS	Adult	Flavor of smokeless tobacco when first started using: Some other flavor - specify
R04_AS1012_OS	Adult	Smokeless tobacco flavor usually/last used: Some other flavor - specify
R04_AS1131_OS	Adult	In past 30 days, smokeless tobacco flavor used: Some other flavor - specify
R04_AU1011_OS	Adult	Flavor of snus when first started using: Some other flavor - specify
R04_AU1012_OS	Adult	Snus flavor usually/last used: Some other flavor - specify
R04_AU1131_OS	Adult	In past 30 days, snus flavor used: Some other flavor - specify
R04_AM0008_RS	Adult	Currently on active duty in the U.S. Armed Forces
R04_AM0050_RS	Adult	Ever served on active duty in the U.S. Armed Forces, Military Reserves or National Guard (Confirmation)
R04_AM0074_RS	Adult	Was on full-time active duty in the U.S. Armed Forces, Military Reserves or National Guard for the entire year in 2014
R04_AM0075_RS	Adult	Lived in the U.S. in any of the 50 states or the District of Columbia during 2014

Table B-7. Wave 4 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R04_CPT05	Parent	Confirm DOB for Wave 1 shadow youth who is expected to be a youth at Wave 3
R04_CPT05C	Parent	Confirm DOB for Wave 1 youth who is expected to be a youth at Wave 3
R04_CPT05D	Parent	Confirm DOB for Wave 1 shadow youth who is expected to be a shadow youth at Wave 3
R04_CPT07	Parent	Corrected Youth DOB
R04_PM0021_NB	Parent	First names of siblings in multiple birth that are identical to youth
R04_PM0030_NB	Parent	First name of sibling that youth is a twin of
R04_PM0035_NB	Parent	First names of siblings that are in the multiple birth with youth
R04_PM0058_NEW_FNAME	Parent	First name of other parental figure or guardian
R04_PM0058_NEW_LNAME	Parent	Last name of other parental figure or guardian
R04_PM0061_NEW_FNAME	Parent	First name of the second other parental figure or guardian
R04_PM0061_NEW_LNAME	Parent	Last name of the second other parental figure or guardian
R04_PT0046_NEW_FNAME	Parent	First name of spouse/partner
R04_PT0046_NEW_LNAME	Parent	Last name of spouse/partner
R04_PARENT_PERSONID	Parent	Wave 4 Parent/guardian Participant ID Number
R04_PM0053	Parent	Confirm parent's relationship to youth
R04_PT0001_OS_1	Parent	Parent or guardian relationship to youth: Other relative
R04_PT0001_OS_2	Parent	Parent or guardian relationship to youth: Other non-relative
R04_PM0052	Parent	Confirm parent's spouse is same as reported in previous wave
R04_PT0046_NEW_AGE	Parent	Age of new spouse/partner
R04_PT0002_OS_1	Parent	Parent's spouse or partner relationship to youth: Other relative - specify
R04_PT0002_OS_2	Parent	Parent's spouse or partner relationship to youth: Non-relative - specify
R04_PM0058_PERSONID	Parent	Other parental figure/guardian participant ID number
R04_PM0058_NEW_AGE	Parent	Age of other parental figure/guardian
R04_PM0061_PERSONID	Parent	Second other parental figure/guardian PID
R04_PM0061_NEW_AGE	Parent	Age of second other parental figure/guardian
R04_PM0059_OS_1	Parent	Other parental figure/guardian's relationship to sampled youth: Other relative - specify
R04_PM0059_OS_2	Parent	Other parental figure/guardian's relationship to sampled youth: Other non-relative - specify
R04_PM0062_OS_1	Parent	Second other parental figure/guardian's relationship to youth: Other relative - specify
R04_PM0062_OS_2	Parent	Second other parental figure/guardian's relationship to youth: Other non-relative - specify
R04_PM0072_OS	Parent	Other language spoken at home: Some other language - specify (parent respondent)
R04_PM0016_NB	Parent	Youth was identified as a twin or part of a multiple birth
R04_PM0017_NB	Parent	Youth and sibling are identical twins
R04_PM0019_NB	Parent	Any sibling in multiple birth identical to youth
R04_PX0186_NB	Parent	Youth has serious difficulty walking or climbing stairs
R04_PX0191_NB	Parent	Youth is blind or has serious difficulty seeing even when wearing glasses
R04_PX0190_NB	Parent	Youth is deaf or has serious difficulty hearing

Table B-7. Wave 4 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R04_LCPT01	Parent	Language in which the CAPI portions of the interview was completed (interviewer report)
R04_E_YOUTH	Youth	Respondent is an emancipated youth
R04_LYH01	Youth	Youth's preferred language to complete ACASI interview
R04_LEY01	Youth	Emancipated Youth's preferred language to complete ACASI interview
R04_YM0007	Youth	Ever served on active duty in the U.S. Armed Forces, Military Reserves or National Guard
R04_YM0011_01	Youth	Branch served when on active duty: Army
R04_YM0011_02	Youth	Branch served when on active duty: Navy
R04_YM0011_03	Youth	Branch served when on active duty: Air Force
R04_YM0011_04	Youth	Branch served when on active duty: Marine Corps
R04_YM0011_05	Youth	Branch served when on active duty: Coast Guard
R04_YM0066	Youth	How well speak English
R04_YM0070	Youth	How well read English
R04_YM0073	Youth	How well write in English
R04_YM0072_OS	Youth	Language other than English spoken at home: Other language - Specify
R04_YC1118_OS	Youth	In past 30 days, how you usually got your own cigarettes - Specify
R04_YC1033_OS	Youth	Retail location where your cigarettes are bought most of the time - Specify
R04_YC1049	Youth	Brand of cigarettes usually/last smoked - Specify
R04_YC1071	Youth	Sub-brand of cigarette product usually/last smoked - Specify
R04_YV9035_OS	Youth	Ever used some other electronic nicotine product - specify
R04_YV1011_OS	Youth	Flavor of first [primary electronic nicotine product] used: Specify.
R04_YV1131_OS	Youth	In past 30 days, used [electronic nicotine products/cartridges/e-liquid] flavored to taste like some other flavor - Specify
R04_YV1118_OS	Youth	In past 30 days, how you usually got your own [electronic nicotine products/cartridges/e-liquid] - Specify
R04_YV1033_OS	Youth	Where [electronic nicotine products/cartridges/e-liquid] is purchased - Specify
R04_YV1049	Youth	Brand of e-liquid usually/last used - Specify
R04_YJ1118TC_OS	Youth	In past 30 days, how you usually got your own traditional cigars as a blunt - Specify
R04_YJ1033TC_OS	Youth	Retail location where your traditional cigars for blunts are bought most of the time - Specify
R04_YJ1049TC	Youth	Brand of traditional cigars usually/last smoked as a blunt - Specify
R04_YJ1071TC	Youth	Sub-brand of traditional cigar product usually/last smoked as a blunt - Specify
R04_YJ1118CG_OS	Youth	In past 30 days, how you usually got your own cigarillo as a blunt - Specify
R04_YJ1033CG_OS	Youth	Retail location where your cigarillos for blunts are bought most of the time - Specify

Table B-7. Wave 4 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R04_YJ1049CG	Youth	Brand of cigarillos usually/last smoked as a blunt - Specify
R04_YJ1071CG	Youth	Sub-brand of cigarillo product usually/last smoked as a blunt - Specify
R04_YJ1118FC_OS	Youth	In past 30 days, how you usually got your own filtered cigars as a blunt - Specify
R04_YJ1033FC_OS	Youth	Retail location where your filtered cigars for blunts are bought most of the time - Specify
R04_YJ1049FC	Youth	Brand of filtered cigars usually/last smoked as a blunt - Specify
R04_YJ1071FC	Youth	Sub-brand of filtered cigar product usually/last smoked as a blunt - Specify
R04_YG1118TC_OS	Youth	In past 30 days, how you usually got your own traditional cigars - Specify
R04_YG1033TC_OS	Youth	Retail location where your traditional cigars are purchased most of the time - Specify
R04_YG1049TC	Youth	Brand of traditional cigars usually/last smoked - Specify
R04_YG1071TC	Youth	Sub-brand of traditional cigar product usually/last smoked - Specify
R04_YG1118CL_OS	Youth	In past 30 days, how you usually got your own cigarillos - Specify
R04_YG1033CL_OS	Youth	Retail location where your cigarillos are purchased most of the time - Specify
R04_YG1049CL	Youth	Brand of cigarillos usually / last smoked - Specify
R04_YG1071CL	Youth	Sub-brand of cigarillo product usually/last smoked - Specify
R04_YG1118FC_OS	Youth	In past 30 days, how you usually got your own filtered cigars - Specify
R04_YG1033FC_OS	Youth	Retail location where your filtered cigars are purchased most of the time - Specify
R04_YG1049FC	Youth	Brand of filtered cigars usually/last smoked - Specify
R04_YG1071FC	Youth	Sub-brand of filtered cigar product usually/last smoked - Specify
R04_YP1118_OS	Youth	In past 30 days, how you usually got your own pipe tobacco - Specify
R04_YP1033_OS	Youth	Retail location where your pipe tobacco are bought most of the time - Specify
R04_YH9011_OS	Youth	Usually smoke a hookah: Somewhere else - Specify
R04_YH1118_OS	Youth	In past 30 days, how you usually got your own shisha or hookah tobacco - Specify
R04_YH1033_OS	Youth	Retail location where your hookah tobacco is purchased most of the time - Specify
R04_YH1049	Youth	Brand of hookah tobacco usually/last smoked - Specify
R04_YH1071	Youth	Sub-brand of hookah tobacco product usually/last smoked - Specify
R04_YH1073	Youth	Smoke hookah because: It is part of my cultural tradition
R04_YU1118_OS	Youth	In past 30 days, how you usually got your own snus pouches - Specify
R04_YU1033_OS	Youth	Retail location where your snus pouches are purchased most of the time - Specify
R04_YU1049	Youth	Brand of snus pouches usually / last used - Specify
R04_YU1071	Youth	Sub-brand of snus pouches usually / last used - Specify

Table B-7. Wave 4 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R04_YS1118_OS	Youth	In past 30 days, how you usually got your own smokeless tobacco - Specify
R04_YS1033_OS	Youth	Retail location where your smokeless tobacco is bought most of the time - Specify
R04_YS1049	Youth	Brand of smokeless tobacco usually/last used - Specify
R04_YS1071	Youth	Sub-brand of smokeless tobacco product usually/last used - Specify
R04_YZ1002_OS	Youth	Ever used any other tobacco products - Specify
R04_YY0601_OS	Youth	First type of tobacco you tried - Specify
R04_YX0686_OS	Youth	In past 12 months, liked or followed brand on social media sites: Other - Specify
R04_YX0136	Youth	Currently pregnant
R04_YX0137_NN	Youth	Number of weeks/months pregnant - Number
R04_YX0137_UN	Youth	Number of weeks/months pregnant - Unit
R04_YX0203_OS	Youth	In past 30 days, has noticed e-cigarettes or other electronic nicotine products being advertised - Specify
R04_YX0677_OS	Youth	In past 30 days, noticed cigarettes or other tobacco products being advertised - Specify
R04_YM0020	Youth	Last grade/year in school completed
R04_YM0021	Youth	Sexual attraction to gender
R04_YM0063	Youth	Sexual orientation
R04_YM0061	Youth	Transgender
R04_YM0062	Youth	Transgender category
R04_LCYS01	Youth	Language in which CAPI portions of youth interview were conducted
R04_YG1011CL_OS	Youth	Flavor of first cigarillo smoked: Some other flavor - specify
R04_YG1011FC_OS	Youth	Flavor of first filtered cigar smoked: Some other flavor - specify
R04_YG1011TC_OS	Youth	Flavor of first traditional cigar smoked: Some other flavor - specify
R04_YG1131CL_OS	Youth	In past 30 days, smoked cigarillos flavored to taste like: Some other flavor - specify
R04_YG1131FC_OS	Youth	In past 30 days, smoked filtered cigars flavored to taste like: Some other flavor - specify
R04_YG1131TC_OS	Youth	In past 30 days, smoked traditional cigars flavored to taste like: Some other flavor - specify
R04_YH1011_OS	Youth	Flavor of first shisha or hookah tobacco smoked: Some other flavor - specify
R04_YH1131_OS	Youth	In past 30 days, smoked hookah tobacco flavored to taste like: Some other flavor - specify
R04_YJ1011CG_OS	Youth	Flavor of first cigarillo smoked as a blunt: Some other flavor - specify
R04_YJ1011FC_OS	Youth	Flavor of first filtered cigar smoked as a blunt: Some other flavor - specify
R04_YJ1011TC_OS	Youth	Flavor of first traditional cigar smoked as a blunt: Some other flavor - specify
R04_YJ1131CG_OS	Youth	In past 30 days, smoked cigarillos as blunts flavored to taste like: Some other flavor - specify
R04_YJ1131FC_OS	Youth	In past 30 days, smoked filtered cigars as blunts flavored to taste like: Some other flavor - specify

Table B-7. Wave 4 questionnaire variables entirely excluded from the PUFs (continued)

Questionnaire variable	Instrument	Questionnaire variable description
R04_YJ1131TC_OS	Youth	In past 30 days, smoked traditional cigars as blunts flavored to taste like: Some other flavor - specify
R04_YS1011_OS	Youth	Flavor of first smokeless tobacco product used: Some other flavor - specify
R04_YS1131_OS	Youth	In past 30 days, used smokeless tobacco flavored to taste like: Some other flavor - specify
R04_YU1011_OS	Youth	Flavor of first snus used: Some other flavor - specify
R04_YU1131_OS	Youth	In past 30 days, used snus flavored to taste like: Some other flavor - specify

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_AM0001 R04_AM0002 R04_AM0003	Adult	R04R_A_AGECAT6	Age range when interviewed (6 levels)
R04_AM0069	Adult	R04R_A_AM0069	Respondent is a citizen of the United States (2 levels)
R04_AM0065	Adult	R04R_A_AM0065_V2	Number of years lived in the United States (2 levels)
R04_AC1007_NB R04_AC1121_NB	Adult	R04R_A_AC1007_NB	Age range when first started smoking cigarettes fairly regularly (6 levels)
R04_AC1020_NB R04_AC1122_NB	Adult	R04R_A_AC1020_NB	Age range when first started smoking cigarettes every day (6 levels)
R04_AV1007_NB R04_AV1121_NB	Adult	R04R_A_AV1007_NB	Age range when first started using [EPRODTYPE1]s fairly regularly (6 levels)
R04_AV1020_NB R04_AV1122_NB	Adult	R04R_A_AV1020_NB	Age range when first started using [EPRODTYPE1]s every day (6 levels)
R04_AG1007TC_NB R04_AG1121TC_NB	Adult	R04R_A_AG1007TC_NB	Age range when first started smoking traditional cigars fairly regularly (6 levels)
R04_AG1020TC_NB R04_AG1122TC_NB	Adult	R04R_A_AG1020TC_NB	Age range when first started smoking cigarillos every day (6 levels)
R04_AG1007CG_NB R04_AG1121CG_NB	Adult	R04R_A_AG1007CG_NB	Age range when first started smoking filtered cigars fairly regularly (6 levels)
R04_AG1020CG_NB R04_AG1122CG_NB	Adult	R04R_A_AG1020CG_NB	Age range when first started smoking cigarillos every day (6 levels)
R04_AG1007FC_NB R04_AG1121FC_NB	Adult	R04R_A_AG1007FC_NB	Age range when first started smoking filtered cigars fairly regularly (6 levels)
R04_AG1020FC_NB R04_AG1122FC_NB	Adult	R04R_A_AG1020FC_NB	Age range when first started smoking filtered cigars every day (6 levels)
R04_AP1007_NB R04_AP1121_NB	Adult	R04R_A_AP1007_NB	Age range when first started smoking a pipe filled with tobacco fairly regularly (6 levels)
R04_AP1020_NB R04_AP1122_NB	Adult	R04R_A_AP1020_NB	Age range when first started smoking a pipe filled with tobacco every day (6 levels)
R04_AH1007_NB R04_AH1121_NB	Adult	R04R_A_AH1007_NB	Age range when first started smoking hookah fairly regularly (6 levels)
R04_AH1020_NB R04_AH1122_NB	Adult	R04R_A_AH1020_NB	Age range when first started smoking hookah every day (6 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_AU1007_NB R04_AU1121_NB	Adult	R04R_A_AU1007_NB	Age range when first started smoking snus pouches fairly regularly (6 levels)
R04_AU1020_NB R04_AU1122_NB	Adult	R04R_A_AU1020_NB	Age range when first started smoking snus pouches every day (6 levels)
R04_AS1007_NB R04_AS1121_NB	Adult	R04R_A_AS1007_NB	Age range when first started using smokeless tobacco fairly regularly (6 levels)
R04_AS1020_NB R04_AS1122_NB	Adult	R04R_A_AS1020_NB	Age range when first started using smokeless tobacco every day (6 levels)
R04_AX0114_NB R04_AX0253_NB	Adult	R04R_A_AX0114_NB	Age range when you were first told you had high blood pressure (6 levels)
R04_AX0115_NB R04_AX0254_NB	Adult	R04R_A_AX0115_NB	Age range when you were first told you had high cholesterol (6 levels)
R04_AX0116_NB R04_AX0255_NB	Adult	R04R_A_AX0116_NB	Age range when you were first told you had congestive heart failure (6 levels)
R04_AX0117_NB R04_AX0256_NB	Adult	R04R_A_AX0117_NB	Age range when you were first told you had a stroke (6 levels)
R04_AX0112_NB R04_AX0252_NB	Adult	R04R_A_AX0112_NB	Age range when you were first told you had a heart attack (6 levels)
R04_AX0120_NB R04_AX0257_NB	Adult	R04R_A_AX0120_NB	Age range when you were first told you had COPD (6 levels)
R04_AX0121_NB R04_AX0258_NB	Adult	R04R_A_AX0121_NB	Age range when you were first told you had chronic bronchitis (6 levels)
R04_AX0123_NB R04_AX0259_NB	Adult	R04R_A_AX0123_NB	Age range when you were first told you had emphysema (6 levels)
R04_AX0124_NB R04_AX0260_NB	Adult	R04R_A_AX0124_NB	Age range when you were first told you had asthma (6 levels)
R04_AX0131_NB R04_AX0261_NB	Adult	R04R_A_AX0131_NB	Age range when you were first told you had gum disease (6 levels)
R04_AX0133_NB R04_AX0262_NB	Adult	R04R_A_AX0133_NB	Age range when you were first told you had pre-cancerous oral lesions (6 levels)
R04_AX0280_NB R04_AX0263_NB	Adult	R04R_A_AX0280_NB	Age range when you were first told you had diabetes, sugar diabetes, high blood sugar or borderline diabetes (6 levels)
R04_AX0143_NB R04_AX0264_NB	Adult	R04R_A_AX0143_NB	Age range when you were first told you had an ulcer (6 levels)
R04_AX0148_NB R04_AX0266_NB	Adult	R04R_A_AX0148_NB	Age range when you were first told you had stomach or gastro-intestinal bleeding (6 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_AX0150_NB R04_AX0267_NB	Adult	R04R_A_AX0150_NB	Age range when you were first told you had osteoporosis (6 levels)
R04_AX0198_NB R04_AX0268_NB	Adult	R04R_A_AX0198_NB	Age range when you were first told you had a bone fracture because you have fragile bones (6 levels)
R04_AX0152_NB R04_AX0269_NB	Adult	R04R_A_AX0152_NB	Age range when you were first told you had a cataract or glaucoma (6 levels)
R04_AX0703 R04_AX0704	Adult	R04R_A_AX0703	Age range when you were first told you had macular degeneration (6 levels)
R04_AX0145_02 R04_AX0145_03 R04_AX0145_04 R04_AX0145_05 R04_AX0145_09 R04_AX0145_12 R04_AX0145_15 R04_AX0145_18 R04_AX0145_19 R04_AX0145_21 R04_AX0145_16 R04_AX0145_23 R04_AX0145_24 R04_AX0145_25 R04_AX0145_27 R04_AX0145_29 R04_AX0145_30 R04_AX0145_31	Adult	R04R_A_AX0145_NONTOB	Type of cancer is a non-tobacco related cancer (2 levels)
R04_AX0145_01 R04_AX0145_06 R04_AX0145_07 R04_AX0145_08 R04_AX0145_10 R04_AX0145_11 R04_AX0145_13 R04_AX0145_14 R04_AX0145_17 R04_AX0145_20 R04_AX0145_22 R04_AX0145_26 R04_AX0145_28	Adult	R04R_A_AX0145_TOB	Type of cancer is a tobacco related cancer (2 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_AX0146_NB_02 R04_AX0265_NB_02 R04_AX0146_NB_03 R04_AX0265_NB_03 R04_AX0146_NB_04 R04_AX0265_NB_04 R04_AX0146_NB_05 R04_AX0265_NB_05 R04_AX0146_NB_09 R04_AX0265_NB_09 R04_AX0146_NB_12 R04_AX0265_NB_12 R04_AX0146_NB_15 R04_AX0265_NB_15 R04_AX0146_NB_18 R04_AX0265_NB_18 R04_AX0146_NB_19 R04_AX0265_NB_19 R04_AX0146_NB_21 R04_AX0265_NB_21 R04_AX0146_NB_16 R04_AX0265_NB_16 R04_AX0146_NB_23 R04_AX0265_NB_23 R04_AX0146_NB_24 R04_AX0265_NB_24 R04_AX0146_NB_25 R04_AX0265_NB_25 R04_AX0146_NB_27 R04_AX0265_NB_27 R04_AX0146_NB_29 R04_AX0265_NB_29 R04_AX0146_NB_30 R04_AX0265_NB_30 R04_AX0146_NB_31 R04_AX0265_NB_31	Adult	R04R_A_AX0146_NONTOB	Age range when nontobacco related cancer was diagnosed (6 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_AX0146_NB_01 R04_AX0265_NB_01 R04_AX0146_NB_06 R04_AX0265_NB_06 R04_AX0146_NB_07 R04_AX0265_NB_07 R04_AX0146_NB_08 R04_AX0265_NB_08 R04_AX0146_NB_10 R04_AX0265_NB_10 R04_AX0146_NB_11 R04_AX0265_NB_11 R04_AX0146_NB_13 R04_AX0265_NB_13 R04_AX0146_NB_14 R04_AX0265_NB_14 R04_AX0146_NB_17 R04_AX0265_NB_17 R04_AX0146_NB_20 R04_AX0265_NB_20 R04_AX0146_NB_22 R04_AX0265_NB_22 R04_AX0146_NB_26 R04_AX0265_NB_26 R04_AX0146_NB_28 R04_AX0265_NB_28	Adult	R04R_A_AX0146_TOB	Age range when tobacco related cancer was diagnosed (6 levels)
R04_AX0135_12M	Adult	R04R_A_AX0135_12M	Outcome of last pregnancy was miscarriage, induced abortion, ectopic or tubal pregnancy or stillbirth (2 levels)
R04_AX0300_12M_01 R04_AX0300_12M_02 R04_AX0300_12M_03 R04_AX0300_12M_04 R04_AX0300_12M_05 R04_AX0300_12M_06 R04_AX0300_12M_07 R04_AX0706	Adult	R04R_A_AX0300_12M	Outcome for live birth was preterm birth, low birth weight, birth defects, placenta previa, placenta abruption, preeclampsia, or cleft lip or palate (2 levels)
R04_AM0018	Adult	R04R_A_AM0018_V2	Highest grade or level of school completed (5 levels)
R04_AM0030	Adult	R04R_A_AM0030	Total household income in the past 12 months (5 levels)
R04_AM0033	Adult	R04R_A_AM0033_V2	Highest grade or year of school completed by mother, step-mother or mother-figure (5 levels)
R04_AM0034	Adult	R04R_A_AM0034_V2	Highest grade or year of school completed by father, step-father or father-figure (5 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_AM0036	Adult	R04R_A_AM0036	In past 12 months, parents' total household income (5 levels)
R04_AM0026_01 R04_AM0026_02 R04_AM0026_03 R04_AM0026_04 R04_AM0026_05 R04_AM0026_06 R04_AM0026_07 R04_AM0026_08	Adult	R04R_A_AM0026_V2	Currently covered by health insurance or health coverage plan (2 levels)
R04_AM0063	Adult	R04R_A_SEXORIENT2	Adult sexual orientation (2 levels)
R04_AX0313 R04_AX0679_FT R04_AX0679_IN R04_AX0316 R04_AX0109 R04_AX0312	Adult	R04R_A_BMI	Body mass index
R04_AT0047	Adult	R04R_A_AT0047	Recoded marital status (3 levels)
R04_AL0040	Adult	R04R_A_AL0040	Indicator of home ownership (2 levels)
R04_AM0017	Adult	R04R_A_AM0017	Recoded reason for not working for pay (5 levels)
R04_AM0042	Adult	R04R_A_AM0042	Recoded where you currently live (3 levels)
R04_AM0072_02 R04_AM0072_03 R04_AM0072_04 R04_AM0072_05 R04_AM0072_06 R04_AM0072_07 R04_AM0072_08 R04_AM0072_09 R04_AM0072_10	Adult	R04R_A_AM0072_02	Recoded other language(s) spoken at home: Chinese, Tagalog, Vietnamese, French, Korean, German, Arabic, Russian or some other language (adult)
R04_AX0066_01 R04_AX0066_02 R04_AX0066_03 R04_AX0066_04 R04_AX0066_05 R04_AX0066_06 R04_AX0066_07 R04_AX0066_08 R04_AX0066_09	Adult	R04R_A_AX0066	Recoded anyone who lives with you now who uses tobacco (4 levels)
R04_AX0093	Adult	R04R_A_AX0093	Extent to which you are able to carry out your everyday physical activities (2 levels)
R04_AX0691	Adult	R04R_A_AX0691	Recoded type of living space currently living in (4 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_AX0757 R04_AX0758	Adult	R04R_A_AX0757	Age range when you were first told by a doctor, therapist or other mental health professional that you had schizophrenia, schizoaffective disorder or psychosis (6 levels)
R04_AX0762 R04_AX0772	Adult	R04R_A_AX0762	Age range when you were first told by a doctor, therapist or other mental health professional that you had a psychotic illness or episode (6 levels)
R04_AC1006_RS R04_AC1120_RS	Adult	R04R_A_AC1006_RS	Age group when first smoked part or all of a cigarette (6 levels)
R04_AG1006TC_RS R04_AG1120TC_RS	Adult	R04R_A_AG1006TC_RS	Age group when first smoked part or all of a traditional cigar, even one or two puffs (6 levels)
R04_AG1006CG_RS R04_AG1120CG_RS	Adult	R04R_A_AG1006CG_RS	Age group when first smoked part or all of a cigarillo, even one or two puffs (6 levels)
R04_AG1006FC_RS R04_AG1120FC_RS	Adult	R04R_A_AG1006FC_RS	Age group when first smoked part or all of a filtered cigar, even one or two puffs (6 levels)
R04_AV1006_RS R04_AV1120_RS	Adult	R04R_A_AV1006_RS	Age group when first used an electronic nicotine product, even one or two times (6 levels)
R04_AU1006_RS R04_AU1120_RS	Adult	R04R_A_AU1006_RS	Age group when first used snus, even one or two times (6 levels)
R04_AP1006_RS R04_AP1120_RS	Adult	R04R_A_AP1006_RS	Age group when you first smoked a pipe filled with tobacco, even one or two puffs (6 levels)
R04_AH1006_RS R04_AH1120_RS	Adult	R04R_A_AH1006_RS	Age group when first smoked tobacco in a hookah, even one or two puffs (6 levels)
R04_AS1006_RS R04_AS1120_RS	Adult	R04R_A_AS1006_RS	Age group when you first used smokeless tobacco, even one or two times (6 levels)
R04_AX0135_RS_01 R04_AX0135_RS_02 R04_AX0135_RS_03 R04_AX0135_RS_04 R04_AX0135_RS_05	Adult	R04R_A_AX0135_RS	Indicator for adverse pregnancy outcomes resulting in no birth (2 levels)
R04_AX0308_RS	Adult	R04R_A_AX0308_RS	Indicator for any pregnancies resulting in a live birth (2 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_AX0300_RS_01 R04_AX0300_RS_02 R04_AX0300_RS_03 R04_AX0300_RS_04 R04_AX0300_RS_05 R04_AX0300_RS_06 R04_AX0300_RS_07	Adult	R04R_A_AX0300_RS	Indicator for adverse pregnancy outcomes resulting in risky birth (2 levels)
R04_AX0086_RS R04_AX0087_RS	Adult	R04R_A_AX0086_RS	Age group when you first drank alcohol at all, counting small tastes or sips (6 levels)
R04_AX0074_RS R04_AX0270_RS	Adult	R04R_A_AX0074_RS	Age group when first alcoholic drink was consumed, other than small tastes or sips (6 levels)
R04_AX0079_RS R04_AX0271_RS	Adult	R04R_A_AX0079_RS	Age group when first used marijuana, hash, THC, grass, pot or weed (6 levels)
R04_AX0082_RS_01 R04_AX0272_RS_01	Adult	R04R_A_AX0082_RS_01	Age range when first used: Ritalin or Adderall (6 levels)
R04_AX0082_RS_02 R04_AX0272_RS_02	Adult	R04R_A_AX0082_RS_02	Age range when first used: Painkillers, sedatives, or tranquilizers (6 levels)
R04_AX0082_RS_03 R04_AX0272_RS_03	Adult	R04R_A_AX0082_RS_03	Age range when first used: Cocaine or crack (6 levels)
R04_AX0082_RS_04 R04_AX0272_RS_04	Adult	R04R_A_AX0082_RS_04	Age range when first used: Stimulants like methamphetamine or speed (6 levels)
R04_AX0082_RS_05 R04_AX0272_RS_05	Adult	R04R_A_AX0082_RS_05	Age range when first used: Other drugs like heroin, inhalants, solvents, or hallucinogens (6 levels)
R04_AM0004_RS	Adult	R04R_A_SEX	Adult respondent sex
R04_AM0005_RS_01 R04_AM0005_RS_02 R04_AM0005_RS_03 R04_AM0005_RS_04 R04_AM0005_RS_05	Adult	R04R_A_HISP	Hispanic origin from the interview (2 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_AM0006_RS_01 R04_AM0006_RS_02 R04_AM0006_RS_03 R04_AM0006_RS_04 R04_AM0006_RS_05 R04_AM0006_RS_06 R04_AM0006_RS_07 R04_AM0006_RS_08 R04_AM0006_RS_09 R04_AM0006_RS_10 R04_AM0006_RS_11 R04_AM0006_RS_12 R04_AM0006_RS_13 R04_AM0006_RS_14	Adult	R04R_A_RACECAT3	Recoded Race from the interview (3 levels)
R04_PT0001	Parent	R04R_Y_PT0001	Recoded parent or guardian relationship to youth (4 levels)
R04_PT0047	Parent	R04R_Y_PT0047	Recoded parent or guardian marital status (3 levels)
R04_PT0045 R04_PM0057 R04_PM0060	Parent	R04R_P_OTHPAR_INHH	Recoded youth has any parental figures/guardians other than the parent/guardian in the house (2 levels)
R04_PT0002	Parent	R04R_Y_PT0002	Recoded parent's spouse or partner relationship to youth (4 levels)
R04_PM0059	Parent	R04R_Y_PM0059	Recoded other parental figure/guardian's relationship to youth (4 levels)
R04_PM0062	Parent	R04R_Y_PM0062	Second other parental figure/guardian's relationship to youth (4 levels)
R04_PT0041_NB R04_PT0253_NB	Parent	R04R_Y_PT0041_NB	Age range youth was first told he/she had high blood pressure (3 levels)
R04_PT0043_NB R04_PT0254_NB	Parent	R04R_Y_PT0043_NB	Age range youth was first told he/she has high cholesterol (3 levels)
R04_PT0038_NB R04_PT0260_NB	Parent	R04R_Y_PT0038_NB	Age range youth was first told he/she has asthma (3 levels)
R04_PT0042_NB R04_PT0263_NB	Parent	R04R_Y_PT0042_NB	Age range youth was first told by a doctor or other health professional that he/she has diabetes, sugar diabetes, high blood sugar or borderline diabetes (3 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_PM0069	Parent	R04R_Y_PM0069	Recoded parent respondent is a citizen of the United States (2 levels)
R04_PM0065 R04_PM0065_NN	Parent	R04R_Y_PM0065_V2	Recoded number of years parents lived in the United States (2 levels)
R04_PM0001	Parent	R04R_Y_PM0001	Recoded highest grade or year of school completed by parent (6 levels)
R04_PM0118	Parent	R04R_Y_PM0118	Recoded highest grade or year of school completed by spouse/guardian (6 levels)
R04_PM0130	Parent	R04R_Y_PM0130	Recoded total household income in past 12 months (5 levels)
R04_PL0040	Parent	R04R_Y_PL0040	Recoded home is owned or rented (2 levels)
R04_PM0072_02 R04_PM0072_03 R04_PM0072_04 R04_PM0072_05 R04_PM0072_06 R04_PM0072_07 R04_PM0072_08 R04_PM0072_09 R04_PM0072_10	Parent	R04R_Y_PM0072_02	Recoded other language(s) spoken at home: Chinese, Tagalog, Vietnamese, French, Korean, German, Arabic, Russian or some other language (parent respondent)
R04_PX0757 R04_PX0758	Parent	R04R_Y_PX0757	Age range when youth was first told by a doctor, therapist or other mental health professional that he/she had schizophrenia, schizoaffective disorder or psychosis (3 levels)
R04_PX0762 R04_PX0772	Parent	R04R_Y_PX0762	Age range when youth was first told by a doctor, therapist or other mental health professional that he/she had a psychotic illness or episode (3 levels)
R04_YM0069	Youth	R04R_Y_YM0069	Recoded citizen of the United States (2 levels)
R04_YM0065	Youth	R04R_Y_YM0065_V2	Recoded number of years youth lived in the United States (2 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_YC1006_NB R04_YC1120_NB	Youth	R04R_Y_YC1006_NB	Age range when first tried cigarette smoking, even one or two puffs (3 levels)
R04_YC1007 R04_YC1121	Youth	R04R_Y_YC1007	Age range when first started smoking cigarettes fairly regularly (3 levels)
R04_YV1006_NB R04_YV1120_NB	Youth	R04R_Y_YV1006_NB	Age range when first tried [primary electronic nicotine product], even one or two times, even one or two puffs (3 levels)
R04_YV1007 R04_YV1121	Youth	R04R_Y_YV1007	Age range when first started using [primary electronic nicotine product]s fairly regularly (3 levels)
R04_YG1006TC_NB R04_YG1120TC_NB	Youth	R04R_Y_YG1006TC_NB	Age range when first tried a traditional cigar, even one or two puffs (3 levels)
R04_YG1007TC R04_YG1121TC	Youth	R04R_Y_YG1007TC	Age range when first started smoking traditional cigars fairly regularly (3 levels)
R04_YG1006CL_NB R04_YG1120CL_NB	Youth	R04R_Y_YG1006CL_NB	Age range when first tried a cigarillo, even one or two puffs (3 levels)
R04_YG1007CL R04_YG1121CL	Youth	R04R_Y_YG1007CL	Age range when first started smoking cigarillos fairly regularly (3 levels)
R04_YG1006FC_NB R04_YG1120FC_NB	Youth	R04R_Y_YG1006FC_NB	Age range when first tried a filtered cigar, even one or two puffs (3 levels)
R04_YG1007FC R04_YG1121FC	Youth	R04R_Y_YG1007FC	Age range when first started smoking filtered cigars fairly regularly (3 levels)
R04_YH1006_NB R04_YH1120_NB	Youth	R04R_Y_YH1006_NB	Age range when first tried smoking hookah, even one or two puffs (3 levels)
R04_YH1007 R04_YH1121	Youth	R04R_Y_YH1007	Age range when first started smoking a hookah fairly regularly (3 levels)
R04_YU1006_NB R04_YU1120_NB	Youth	R04R_Y_YU1006_NB	Age range when first tried snus pouches, even one or two times (3 levels)
R04_YU1007 R04_YU1121	Youth	R04R_Y_YU1007	Age range when first started using snus pouches fairly regularly (3 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_YS1006_NB R04_YS1120_NB	Youth	R04R_Y_YS1006_NB	Age range when first tried smokeless tobacco, even one or two times (3 levels)
R04_YS1007 R04_YS1121	Youth	R04R_Y_YS1007	Age range when first started using smokeless tobacco fairly regularly (3 levels)
R04_YX0671_01 R04_YX0671_02 R04_YX0671_03 R04_YX0671_04 R04_YX0671_05 R04_YX0671_06 R04_YX0671_07 R04_YX0671_08	Youth	R04R_Y_YX0671	Recoded anyone who lives with you now use tobacco (4 levels)
R04_YT0038_NB R04_YT0260_NB	Youth	R04R_Y_YT0038_NB	Age range when you were first told you had asthma (3 levels)
R04_YX0086_NB R04_YX0087_NB	Youth	R04R_Y_YX0086_NB	Age range when first drank alcohol at all, counting small tastes or sips (3 levels)
R04_YX0074_NB R04_YX0270_NB	Youth	R04R_Y_YX0074_NB	Age range when had first alcoholic drink, other than small tastes or sips (3 levels)
R04_YX0079_NB R04_YX0271_NB	Youth	R04R_Y_YX0079_NB	Age range when first used marijuana, hash, THC or grass (3 levels)
R04_YX0082_NB_01 R04_YX0272_NB_01	Youth	R04R_Y_YX0082_NB_01	Age range when first used: Ritalin or Adderall (3 levels)
R04_YX0082_NB_02 R04_YX0272_NB_02	Youth	R04R_Y_YX0082_NB_02	Age range when first used: Painkillers, sedatives or tranquilizers (3 levels)
R04_YX0082_NB_03 R04_YX0272_NB_03	Youth	R04R_Y_YX0082_NB_03	Age range when first used: Cocaine or crack (3 levels)
R04_YX0082_NB_04 R04_YX0272_NB_04	Youth	R04R_Y_YX0082_NB_04	Age range when first used: Stimulants like methamphetamine or speed (3 levels)
R04_YX0082_NB_05 R04_YX0272_NB_05	Youth	R04R_Y_YX0082_NB_05	Age range when first used: Any other drugs like heroin, inhalants, solvents or hallucinogens (3 levels)
R04_YM0019 R04_YM0018	Youth	R04R_Y_YM0018	Recoded grade level (If on holiday or break - grade level entering when returning to school) (7 levels)

Table B-8. Wave 4 questionnaire variables excluded from the PUFs that are replaced with a derived variable (continued)

Questionnaire Variable	Instrument	PUF Derived Variable	PUF Derived Variable Description
R04_YM0004_NB	Youth	R04R_Y_SEX	Youth respondent sex
R04_YM0005_NB_01 R04_YM0005_NB_02 R04_YM0005_NB_03 R04_YM0005_NB_04 R04_YM0005_NB_05	Youth	R04R_Y_HISP	DERIVED - Hispanic origin from the interview (2 levels)
R04_YM0006_NB_01 R04_YM0006_NB_02 R04_YM0006_NB_03 R04_YM0006_NB_04 R04_YM0006_NB_05 R04_YM0006_NB_06 R04_YM0006_NB_07 R04_YM0006_NB_08 R04_YM0006_NB_09 R04_YM0006_NB_10 R04_YM0006_NB_11 R04_YM0006_NB_12 R04_YM0006_NB_13 R04_YM0006_NB_14	Youth	R04R_Y_RACECAT3	Recoded Race from the interview (3 levels)
R04_PT0007_FT R04_PT0007_IN R04_PT0007_MT R04_PT0008_LB R04_PT0008_KG R04_YX0310 R04_YT0007_FT R04_YT0007_IN R04_YX0311 R04_YT0008 R04_YX0312	Youth/ Parent	R04R_Y_BMI	Body mass index
R04_YM0072_02 R04_YM0072_03 R04_YM0072_04 R04_YM0072_05 R04_YM0072_06 R04_YM0072_07 R04_YM0072_08 R04_YM0072_09 R04_YM0072_10	Youth	R04R_Y_YM0072_02	Recoded other language(s) spoken at home: Chinese, Tagalog, Vietnamese, French, Korean, German, Arabic, Russian or some other language (youth)

Appendix C
Variables with Coded Outlier Values

Table C-1. Wave 1 questionnaire variables with coded outlier values

Variable name	Data file	Variable description
R01_AC1006	Adult	Age when first smoked part or all of a cigarette
R01_AC1007	Adult	Age when first started smoking cigarettes fairly regularly
R01_AC1020	Adult	Age when first started smoking cigarettes every day
R01_AC1009_NN	Adult	How long since you completely quit smoking cigarettes - Number
R01_AC9002_YR	Adult	How long smoking / smoked cigarettes fairly regularly - Years
R01_AC9002_MO	Adult	How long smoking / smoked cigarettes fairly regularly - Months
R01_AC9003_NN	Adult	How many cigarettes smoked per day when you smoked fairly regularly - Number
R01_AC1021_NN	Adult	Average number of cigarettes now smoked each day - Number
R01_AC1022	Adult	Number of days smoked cigarettes in past 30 days
R01_AC1023_NN	Adult	Average number of cigarettes smoked per day on days smoked in past 30 days - Number
R01_AC9005_NN	Adult	Average number of cigarettes smoked per day when smoked fairly regularly in the past - Number
R01_AC1024_NN	Adult	Time to first cigarette after waking - Number
R01_AC9006_NN	Adult	Average number of cigarettes smoked per day 12 months ago - Number
R01_AC1051RY_NN	Adult	How long been smoking regular brand of roll-your-own cigarette tobacco - Number
R01_AC1041_D	Adult	Amount usually paid for a carton of cigarettes - Dollars
R01_AC1041_C	Adult	Amount usually paid for a carton of cigarettes - Cents
R01_AC1042_D	Adult	Amount usually paid for a pack of cigarettes - Dollars
R01_AC1042_C	Adult	Amount usually paid for a pack of cigarettes - Cents
R01_AC1043_D	Adult	Amount usually paid for a single cigarette - Dollars
R01_AC1043_C	Adult	Amount usually paid for a single cigarette - Cents
R01_AC1051MC_NN	Adult	How long smoked regular brand of cigarettes - Number
R01_AE1006	Adult	Age when first time used an e-cigarette, even one or two times
R01_AE1009_NN	Adult	How long since you took last drag from an e-cigarette - Number
R01_AE9029_NN	Adult	Time since you took last puff from an e-cigarette - Number
R01_AE1024_NN	Adult	Time to first e-cigarette puff after waking - Number
R01_AE1051_NN	Adult	How long used regular brand of e-cigarettes - Number
R01_AG1006TC	Adult	Age when first smoked part or all of a traditional cigar, even one or two puffs
R01_AG1007TC	Adult	Age when first started smoking traditional cigars fairly regularly
R01_AG1009TC_NN	Adult	How long since you last smoked a traditional cigar - Number
R01_AG1022TC	Adult	Number of days smoked traditional cigars in past 30 days
R01_AG9029TC_NN	Adult	Time since you last smoked traditional cigars - Number
R01_AG1024TC_NN	Adult	Time to first traditional cigar after waking on days smoked - Number
R01_AG1098TC	Adult	Number of times used a coupon to buy traditional cigars in past 30 days
R01_AG1051TC_NN	Adult	How long smoked regular brand of traditional cigar - Number

Table C-1. Wave 1 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R01_AG1006CG	Adult	Age when first time smoked part or all of a cigarillo, even one or two puffs
R01_AG1007CG	Adult	Age when first started smoking cigarillos fairly regularly
R01_AG1009CG_NN	Adult	How long since you last smoked a cigarillo - Number
R01_AG1022CG	Adult	Number of days smoked cigarillos in past 30 days
R01_AG9029CG_NN	Adult	Time since you last smoked cigarillos - Number
R01_AG1024CG_NN	Adult	Time to first cigarillo after waking on days smoked - Number
R01_AG1051CG_NN	Adult	How long smoked regular brand of cigarillo - Number
R01_AG1006FC	Adult	Age when first smoked part or all of a filtered cigar, even one or two puffs
R01_AG1007FC	Adult	Age when first started smoking filtered cigars fairly regularly
R01_AG1020FC	Adult	Age when first started smoking filtered cigars every day
R01_AG1009FC_NN	Adult	How long since you last smoked a filtered cigar - Number
R01_AG1022FC	Adult	Number of days smoked filtered cigars in past 30 days
R01_AG9029FC_NN	Adult	Time since you last smoked filtered cigars - Number
R01_AG1024FC_NN	Adult	Time to first filtered cigar after waking on days smoked - Number
R01_AG1051FC_NN	Adult	How long smoked regular brand of filtered cigar - Number
R01_AP1006	Adult	Age when first smoked part or all of a pipe filled with tobacco, even one or two puffs
R01_AP1007	Adult	Age when first started smoking a pipe filled with tobacco fairly regularly
R01_AP1009_NN	Adult	How long since you last smoked a pipe filled with tobacco - Number
R01_AP1022	Adult	Number of days smoked a pipe filled with tobacco in past 30 days
R01_AP9029_NN	Adult	Time since you last smoked a pipe filled with tobacco - Number
R01_AP1024_NN	Adult	Time to first pipe bowl after waking on days smoked - Number
R01_AH1006	Adult	Age when first smoked hookah, even one or two puffs
R01_AH1007	Adult	Age when first started smoking hookah fairly regularly
R01_AH1009_NN	Adult	How long since you last smoked a hookah - Number
R01_AH1024_NN	Adult	Time to first puff from hookah after waking on days smoked - Number
R01_AH1051_NN	Adult	How long used regular brand of hookah tobacco - Number
R01_AS1006SU	Adult	Age when first used snus pouches, even one or two times
R01_AS1009SU_NN	Adult	How long since you last used snus pouches - Number
R01_AS9029SU_NN	Adult	Time since you last used snus pouches - Number
R01_AS1024SU_NN	Adult	Time to first snus pouch after waking on days used - Number
R01_AS1051SU_NN	Adult	How long used regular brand of snus pouches - Number
R01_AS1006SM	Adult	Age when first used smokeless tobacco, even one or two times
R01_AS1007SM	Adult	Age when first started using smokeless tobacco fairly regularly
R01_AS1009SM_NN	Adult	How long since you last used smokeless tobacco - Number
R01_AS9029SM_NN	Adult	Time since you last used smokeless tobacco - Number
R01_AS1024SM_NN	Adult	Time to first use of smokeless tobacco on days used - Number
R01_AS1051SM_NN	Adult	How long used regular brand of smokeless tobacco - Number
R01_AD1006	Adult	Age when first used dissolvable tobacco, even one or two times
R01_AD1009_NN	Adult	How long since you last used dissolvable tobacco - Number
R01_AD9029_NN	Adult	Time since you last used dissolvable tobacco - Number
R01_AD1051_NN	Adult	How long used regular brand of dissolvable tobacco - Number
R01_AY0010_NN	Adult	Time to first use of any tobacco product after waking - Number

Table C-1. Wave 1 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R01_AN0130_NN	Adult	Length of time you stopped smoking / using tobacco product(s) because you were trying to quit, in the past 12 months - Number
R01_AN0251_NN	Adult	How long been using nicotine medication to help quit - Number
R01_AN0309	Adult	Number of nicotine patches, gum, inhaler, nasal spray, lozenges or pills used today / yesterday / the day before yesterday
R01_AN0175_NN	Adult	How long ago did you stop using nicotine medication - Number
R01_AN0252_NN	Adult	How long have you been using a prescription drug to help quit - Number
R01_AN0205_NN	Adult	How long ago you stopped using prescription medication - Number
R01_AN0130E_NN	Adult	Length of time you stopped smoking / using tobacco product(s) because you were trying to quit, in the past 12 months - Number
R01_AX0068	Adult	Number of hours in past 7 days that you were in close contact with others when they were smoking
R01_AX0155	Adult	Number of times you visited an emergency room or urgent care center for a health problem of your own in past 12 months
R01_AX0114	Adult	Age when you were first told you had high blood pressure
R01_AX0115	Adult	Age when you were first told you had high cholesterol
R01_AX0116	Adult	Age when you were first told you had congestive heart failure
R01_AX0117	Adult	Age when you were first told you had a stroke
R01_AX0112	Adult	Age when you were first told you had a heart attack
R01_AX0120	Adult	Age when you were first told you had COPD
R01_AX0121	Adult	Age when you were first told you had chronic bronchitis
R01_AX0123	Adult	Age when you were first told you had emphysema
R01_AX0124	Adult	Age when you were first told you had asthma
R01_AX0131	Adult	Age when you were first told you had gum disease
R01_AX0133	Adult	Age when you were first told you had pre-cancerous oral lesions
R01_AX0280	Adult	Age when you were first told you had diabetes, sugar diabetes, high blood sugar or borderline diabetes
R01_AX0143	Adult	Age when you were first told you had an ulcer
R01_AX0148	Adult	Age when you were first told you had stomach or gastro-intestinal bleeding
R01_AX0150	Adult	Age when you were first told you had osteoporosis
R01_AX0198	Adult	Age when you were first told you had a bone fracture because you have fragile bones
R01_AX0152	Adult	Age when you were first told you had a cataract or glaucoma
R01_AX0146_01	Adult	Age when bladder cancer was first diagnosed
R01_AX0146_04	Adult	Age when brain cancer was first diagnosed
R01_AX0146_06	Adult	Age when cervix (cervical) cancer was first diagnosed
R01_AX0146_07	Adult	Age when colon cancer was first diagnosed
R01_AX0146_09	Adult	Age when gallbladder cancer was first diagnosed
R01_AX0146_10	Adult	Age when kidney cancer was first diagnosed
R01_AX0146_16	Adult	Age when melanoma was first diagnosed
R01_AX0146_17	Adult	Age when mouth / tongue / lip cancer was first diagnosed
R01_AX0146_21	Adult	Age when prostate cancer was first diagnosed
R01_AX0146_24	Adult	Age when skin (unknown kind) cancer was first diagnosed
R01_AX0146_30	Adult	Age when uterus (uterine) cancer was first diagnosed
R01_AX0086	Adult	Age when first drank alcohol at all, counting small tastes or sips
R01_AX0074	Adult	Age when first alcoholic drink was consumed

Table C-1. Wave 1 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R01_AX0079	Adult	Age when first used marijuana, hash, THC or grass
R01_AX0082_01	Adult	Age when first started using: Ritalin or Adderall
R01_AX0082_02	Adult	Age when first started using: Painkillers, sedatives or tranquilizers
R01_AX0082_03	Adult	Age when first started using: Cocaine or crack
R01_AX0082_04	Adult	Age when first started using: Stimulants like methamphetamine or speed
R01_AX0082_05	Adult	Age when first started using: Any other drugs like heroin, inhalants, solvents, or hallucinogens
R01_AX0137_NN	Adult	Weeks / Months pregnant - Number
R01_AX0309	Adult	Calendar year of most recent pregnancy
R01_AN0145_NN	Adult	Longest time period for which you stopped smoking / using tobacco product(s) because you were trying to quit, in the past 12 months - Number
R01_AN0145E_NN	Adult	Longest time period for which you stopped using e-cigarettes because you were trying to quit, in the past 12 months - Number
R01_AN0165_NN	Adult	Length of time used nicotine medication during last tobacco product(s) quit attempt - Number
R01_AN0165E_NN	Adult	Length of time used nicotine medication during last e-cigarette quit attempt - Number
R01_AN0195_NN	Adult	Length of time used prescription medication during last tobacco product(s) quit attempt - Number
R01_AN0195E_NN	Adult	Length of time used prescription medication during last e-cigarette quit attempt - Number
R01_AN0135_VALUE_INFO	Adult	Type of response provided for R01_AN0135
R01_AN0135E_VALUE_INFO	Adult	Type of response provided for R01_AN0135E
R01_AN0115	Adult	Number of times tried to quit smoking / using tobacco product(s) in past 12 months
R01_PT0007_FT	Youth	Youth's current height (feet)
R01_PT0007_IN	Youth	Youth's current height (inches)
R01_PT0008_LB	Youth	Youth's current weight (pounds)
R01_PT0043	Youth	Age youth was first told he/she has high cholesterol
R01_PT0038	Youth	Age youth was first told he/she has asthma
R01_PT0042	Youth	Age youth was first told by a doctor or other health professional that he/she has diabetes, sugar diabetes, high blood sugar or borderline diabetes
R01_YC1006	Youth	Age when first tried cigarette smoking, even one or two puffs
R01_YC1022	Youth	Number of days smoked cigarettes in past 30 days
R01_YC1051_NN	Youth	How long smoked regular brand of cigarettes - Number
R01_YE1006	Youth	Age when first tried an e-cigarette, even one or two times
R01_YE1022	Youth	Number of days used an e-cigarette in past 30 days
R01_YG1006TC	Youth	Age when first tried a traditional cigar, even one or two puffs
R01_YG1022TC	Youth	Number of days smoked a traditional cigar in past 30 days
R01_YG1006CL	Youth	Age when first tried a cigarillo, even one or two puffs
R01_YG1022CL	Youth	Number of days smoked a cigarillo in past 30 days
R01_YG1006FC	Youth	Age when first tried a filtered cigar, even one or two puffs
R01_YP1006	Youth	Age when first tried pipe tobacco, even one or two puffs
R01_YH1006	Youth	Age when first tried smoking a hookah, even one or two puffs

Table C-1. Wave 1 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R01_YS1006SU	Youth	Age when first tried snus pouches, even one or two times
R01_YD1006	Youth	Age when first tried a dissolvable tobacco product, even one or two times
R01_YB1006BD	Youth	Age when first tried a bidi, even one or two puffs
R01_YB1006KK	Youth	Age when first tried a kretek, even one or two puffs
R01_YX0086	Youth	Age when first drank alcohol at all
R01_YX0074	Youth	Age when consumed first alcoholic drink
R01_YX0079	Youth	Age when first used marijuana, hash, THC or grass
R01_YX0082_01	Youth	Age when first used: Ritalin or Adderall
R01_YX0082_02	Youth	Age when first used: Painkillers, sedatives or tranquilizers
R01_YX0082_03	Youth	Age when first used: Cocaine or crack
R01_YX0082_04	Youth	Age when first used: Stimulants like methamphetamine or speed
R01_YX0082_05	Youth	Age when first used: Any other drugs like heroin, inhalants, solvents or hallucinogens

Table C-2. Wave 2 questionnaire variables with coded outlier values

Variable name	Data file	Variable description
R02_AC1009_NN	Adult	How long since last smoked a cigarette - Number
R02_AO0103C	Adult	Number of puffs from e-cigarette today/yesterday/day before yesterday
R02_AO0103H	Adult	Number of puffs from e-hookah today/yesterday/day before yesterday
R02_AE1009_NN	Adult	How long since you took last puff from an e-cigarette - Number
R02_AG1009TC_NN	Adult	How long since last smoked a traditional cigar - Number
R02_AG1009CG_NN	Adult	How long since last smoked a cigarillo - Number
R02_AG1009FC_NN	Adult	How long since last smoked a filtered cigar - Number
R02_AG1009TJ_NN	Adult	How long since last smoked a traditional cigar as a blunt - Number
R02_AG1009GJ_NN	Adult	How long since last smoked a cigarillo as a blunt - Number
R02_AG1009FJ_NN	Adult	How long since last smoked a filtered cigar as a blunt - Number
R02_AP1009_NN	Adult	How long since you last smoked a pipe filled with tobacco - Number
R02_AH1009_NN	Adult	How long since you last smoked a hookah - Number
R02_AS1009SU_NN	Adult	How long since you last used snus pouches - Number
R02_AS1009SM_NN	Adult	How long since you last used smokeless tobacco - Number
R02_AD1009_NN	Adult	How long since you last used dissolvable tobacco - Number
R02_AC1023_NN	Adult	In past 30 days, average number of cigarettes smoked per day on days smoked - Number
R02_AC1024_NN	Adult	Time to first cigarette after waking on days smoked - Number
R02_AC1051MC_NN	Adult	How long smoked regular brand of manufactured cigarettes - Number
R02_AE9029_NN	Adult	Time since last puff from an e-cigarette - Number
R02_AE1024_NN	Adult	Time to first e-cigarette puff after waking on days used - Number
R02_AE1051_NN	Adult	How long used regular brand of e-cigarette/e-cigarette cartridge/e-liquid - Number
R02_AO9029_NN	Adult	Time since last puff from an electronic nicotine product (other than e-cigarettes) - Number
R02_AG9029TC_NN	Adult	Time since last smoked traditional cigars - Number
R02_AG1024TC_NN	Adult	Time to first traditional cigar after waking on days smoked - Number
R02_AG1051TC_NN	Adult	How long smoked regular brand of traditional cigar - Number
R02_AG9029CG_NN	Adult	Time since last smoked cigarillos - Number
R02_AG9029FC_NN	Adult	Time since last smoked filtered cigars - Number
R02_AG1024FC_NN	Adult	Time to first filtered cigar after waking on days smoked - Number
R02_AP9029_NN	Adult	Time since last smoked a pipe filled with tobacco - Number
R02_AH1024_NN	Adult	Time to first puff from hookah after waking on days smoked - Number
R02_AS9029SU_NN	Adult	Time since last used snus pouches - Number
R02_AS1024SU_NN	Adult	Time to first snus pouch after waking on days used - Number
R02_AS9029SM_NN	Adult	Time since last used smokeless tobacco - Number
R02_AS1024SM_NN	Adult	Time to first use of smokeless tobacco after waking on days used - Number
R02_AD9029_NN	Adult	Time since last used dissolvable tobacco - Number
R02_AY0010_NN	Adult	Time to first use of any tobacco product after waking - Number

Table C-2. Wave 2 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R02_AN0130_NN	Adult	In past 12 months, length of time you stopped smoking/using tobacco product(s) because you were trying to quit - Number
R02_AN0135_VALUE_INFO	Adult	Type of response provided for R02_AN0135
R02_AN0145_NN	Adult	In past 12 months, longest time period for which you stopped smoking/using tobacco product(s) because you were trying to quit - Number
R02_AN0251_NN	Adult	How long been using nicotine medication to help quit - Number
R02_AN0175_NN	Adult	How long ago did you stop using nicotine medication - Number
R02_AN0195_NN	Adult	Length of time used prescription medication during last tobacco product(s) quit attempt - Number
R02_AN0252_NN	Adult	How long have you been using a prescription drug to help quit - Number
R02_AN0205_NN	Adult	How long ago you stopped using prescription medication - Number
R02_AN0130E_NN	Adult	In past 12 months, length of time you stopped using e-cigarettes because you were trying to quit - Number
R02_AN0135E_VALUE_INFO	Adult	Type of response provided for R02_AN0135E
R02_AN0145E_NN	Adult	In past 12 months, longest time period for which you stopped using e-cigarettes because you were trying to quit - Number
R02_AN0165E_NN	Adult	Length of time used nicotine medication during last e-cigarette quit attempt - Number
R02_AN0195E_NN	Adult	Length of time used prescription medication during last e-cigarette quit attempt - Number
R02_AX0068	Adult	In past 7 days, Number of hours that you were in close contact with others when they were smoking
R02_AX0679_FT	Adult	Height without shoes: Feet
R02_AX0679_IN	Adult	Height without shoes: Inches
R02_AX0316	Adult	Height without shoes: Meters
R02_AX0109	Adult	Current weight: Pounds
R02_AX0312	Adult	Current weight: Kilograms
R02_AX0242_NN	Adult	On days when you do physical activity or exercise of at least moderate intensity, how long do you do these activities - Number
R02_AX0137_NN	Adult	Weeks/Months pregnant - Number
R02_PT0007_FT	Youth	Youth's current height (feet)
R02_PT0007_IN	Youth	Youth's current height (inches)
R02_PT0007_MT	Youth	Youth's current height (meters)
R02_PT0008_LB	Youth	Youth's current weight (pounds)
R02_YE1007	Youth	Age when first started using e-cigarettes fairly regularly
R02_YE1041_D	Youth	Amount paid for last bottle or container of e-liquid - Dollars
R02_YE1041_C	Youth	Amount paid for last bottle or container of e-liquid - Cents
R02_YG1007TC	Youth	Age when first started smoking traditional cigars fairly regularly
R02_YG1034GJ_D	Youth	Amount paid when last time bought a box of cigarillos for blunts/pack of cigarillos for blunts/single cigarillo for blunts - Dollars
R02_YS1007SM	Youth	Age when first started using smokeless tobacco fairly regularly
R02_YT0007_FT	Youth	Current height: Feet

Table C-2. Wave 2 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R02_YT0007_IN	Youth	Current height: Inches
R02_YX0311	Youth	Current height: Meters
R02_YT0008	Youth	Current weight: Pounds
R02_YX0312	Youth	Current weight: Kilograms
R02_YT0038_NB	Youth	Age when you were first told you had asthma
R02_YX0137_NN	Youth	Number of weeks/months pregnant - Number
R02_YX0086_NB	Youth	Age when first drank alcohol at all
R02_YX0074_NB	Youth	Age when consumed first alcoholic drink

Table C-3. Wave 3 questionnaire variables with coded outlier values

Variable name	Data file	Variable description
R03_AM0065	Adult	Number of years lived in the United States
R03_AC1009_NN	Adult	How long since last smoked a cigarette - Number
R03_AV1009EC_NN	Adult	How long since last took a puff from an e-cigarette - Number
R03_AV1009EG_NN	Adult	How long since last took a puff from an e-cigar - Number
R03_AV1009EP_NN	Adult	How long since last took a puff from an e-pipe - Number
R03_AV1009EH_NN	Adult	How long since last took a puff from an e-hookah - Number
R03_AV1009EN_NN	Adult	How long since last took a puff from an electronic nicotine product - Number
R03_AG1009TC_NN	Adult	How long since last smoked a traditional cigar - Number
R03_AG1009CG_NN	Adult	How long since last smoked a cigarillo - Number
R03_AG1009FC_NN	Adult	How long since last smoked a filtered cigar - Number
R03_AG1009TJ_NN	Adult	How long since last smoked a traditional cigar as a blunt - Number
R03_AG1009GJ_NN	Adult	How long since last smoked a cigarillo as a blunt - Number
R03_AG1009FJ_NN	Adult	How long since last smoked a filtered cigar as a blunt - Number
R03_AP1009_NN	Adult	How long since you last smoked a pipe filled with tobacco - Number
R03_AH1009_NN	Adult	How long since you last smoked a hookah - Number
R03_AU1009_NN	Adult	How long since you last used snus pouches - Number
R03_AS1009_NN	Adult	How long since you last used smokeless tobacco - Number
R03_AC1020_NB	Adult	Age when first started smoking cigarettes every day
R03_AC1024_NN	Adult	Time to first cigarette after waking on days smoked - Number
R03_AC1024_30D_NN	Adult	Time to first cigarette after waking on days smoked - Number
R03_AC0103MC	Adult	Number of cigarettes smoked today/yesterday/day before yesterday
R03_AC1040	Adult	Number of cigarette packs in the carton you usually buy
R03_AC1051MC_NN	Adult	How long smoked regular brand of manufactured cigarettes - Number
R03_AV1007_NB	Adult	Age when first started using [primary electronic nicotine product]s fairly regularly
R03_AV9045	Adult	Number of milliliters of e-liquid your tank system holds/held
R03_AM0103	Adult	How many puffs of marijuana from a [primary electronic nicotine product] have/did you taken/take
R03_AV9029_NN	Adult	Time since you took last puff from an [primary electronic nicotine product] - Number
R03_AV1024_NN	Adult	Time to first [primary electronic nicotine product] after waking - Number
R03_AV1024_30D_NN	Adult	Time to first [primary electronic nicotine product]s after waking - Number
R03_AV0103	Adult	Number of puffs taken from [primary electronic nicotine product] today/yesterday/day before yesterday
R03_AV1051_NN	Adult	How long used regular brand of [electronic nicotine products/cartridges/e-liquid] - Number

Table C-3. Wave 3 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R03_AV2029_NN	Adult	How long since last took a puff from [secondary electronic nicotine product] - Number
R03_AJ9029TC_NN	Adult	Time since last smoked traditional cigar as blunt - Number
R03_AJ0103TC	Adult	Number of traditional cigars as a blunt smoked today/yesterday/day before yesterday
R03_AG9029TC_NN	Adult	Time since last smoked traditional cigars - Number
R03_AG1024TC_NN	Adult	Time to first traditional cigar after waking on days smoked - Number
R03_AG1024TC_30D_NN	Adult	Time to first traditional cigar after waking on days smoked - Number
R03_AG1040TC	Adult	Number of traditional cigars that come/came in box or pack usually buy/bought
R03_AG1007CG_NB	Adult	Age when first started smoking cigarillos fairly regularly
R03_AG9029CG_NN	Adult	Time since last smoked cigarillos - Number
R03_AG1024CG_NN	Adult	Time to first cigarillo after waking on days smoked - Number
R03_AG1024CG_30D_NN	Adult	Time to first cigarillo after waking on days smoked - Number
R03_AG1007FC_NB	Adult	Age when first started smoking filtered cigars fairly regularly
R03_AG9029FC_NN	Adult	Time since last smoked filtered cigars - Number
R03_AG1024FC_NN	Adult	Time to first filtered cigar after waking on days smoked - Number
R03_AG1024FC_30D_NN	Adult	Time to first filtered cigar after waking on days smoked - Number
R03_AP1024_NN	Adult	Time to first pipe bowl after waking on days smoked - Number
R03_AP1024_30D_NN	Adult	Time to first pipe bowl after waking on days smoked - Number
R03_AU1024_30D_NN	Adult	Time to first snus pouch of the day after waking on the days used - Number
R03_AU9002	Adult	Number of days to use up the container of snus
R03_AS1007_NB	Adult	Age when first started using smokeless tobacco fairly regularly
R03_AS9029_NN	Adult	Time since last used smokeless tobacco - Number
R03_AS1024_NN	Adult	Time to first use of smokeless tobacco after waking on days used - Number
R03_AY0010_NN	Adult	Time to first use of any tobacco product after waking - Number
R03_AN0130E_NN	Adult	In past 12 months, length of time you stopped using [e-cigarettes/electronic nicotine products] because you were trying to quit - Number
R03_AN0135E_VALUE_INFO	Adult	Type of response provided for R03_AN0135E
R03_AN0145E_NN	Adult	In past 12 months, longest time period for which you stopped using [e-cigarettes/electronic nicotine products] because you were trying to quit - Number
R03_AN0130_NN	Adult	In past 12 months, length of time you stopped smoking/using tobacco product(s) because you were trying to quit - Number

Table C-3. Wave 3 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R03_AN0135_VALUE_INFO	Adult	Type of response provided for R03_AN0135
R03_AN0145_NN	Adult	In past 12 months, longest time period for which you stopped smoking/using tobacco product(s) because you were trying to quit - Number
R03_AN0251_NN	Adult	How long been using NRT - Number
R03_AN0205_NN	Adult	How long ago you stopped using prescription medication - Number
R03_AX0068	Adult	In past 7 days, number of hours that you were in close contact with others when they were smoking
R03_AX0679_FT	Adult	Height without shoes: Feet
R03_AX0679_IN	Adult	Height without shoes: Inches
R03_AX0316	Adult	Height without shoes: Meters
R03_AX0109	Adult	Current weight: Pounds
R03_AX0312	Adult	Current weight: Kilograms
R03_AX0242_NN	Adult	On days when you do physical activity or exercise of at least moderate intensity, how long do you do these activities - Number
R03_AX0155	Adult	In past 12 months, number of times you visited an emergency room or urgent care center for a health problem of your own
R03_AX0117_NB	Adult	Age when you were first told you had a stroke
R03_AX0112_NB	Adult	Age when you were first told you had a heart attack or needed bypass surgery
R03_AX0695	Adult	Number of permanent teeth removed because of tooth decay or gum disease
R03_AX0729_NN	Adult	Time since you had your teeth cleaned by a dentist, hygienist, or other health professional - Number
R03_AX0143_NB	Adult	Age when you were first told you had an ulcer
R03_AX0148_NB	Adult	Age when you were first told you had stomach or gastro-intestinal bleeding
R03_AX0150_NB	Adult	Age when you were first told you had osteoporosis
R03_AX0198_NB	Adult	Age when you were first told you had a bone fracture because you have fragile bones
R03_AX0703	Adult	Age when first told you had macular degeneration
R03_AX0137_NN	Adult	Weeks/Months pregnant - Number
R03_PT0007_MT	Youth	Youth's current height (meters)
R03_PT0008_LB	Youth	Youth's current weight (pounds)
R03_PM0065	Youth	Number of years lived in the United States
R03_YM0065	Youth	Number of years lived in the United States
R03_YV9045	Youth	Milliliters of e-liquid tank holds
R03_YV9040	Youth	In past 30 days, number of [electronic nicotine products/cartridges/milliliters of e-liquid] used per day on days used [primary electronic nicotine product]
R03_YV2029_NN	Youth	How long since you took a puff from [secondary electronic nicotine product] - Number

Table C-3. Wave 3 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R03_YC1220	Youth	In past 12 months, number of times you stopped smoking for one day or longer because you were trying to quit smoking cigarettes for good
R03_YX0068	Youth	In past seven days, number of hours that you were in close contact with others when they were smoking
R03_YT0007_FT	Youth	Current height: Feet
R03_YT0007_IN	Youth	Current height: Inches
R03_YX0311	Youth	Current height: Meters
R03_YT0008	Youth	Current weight: Pounds
R03_YX0312	Youth	Current weight: Kilograms

Table C-4. Wave 4 questionnaire variables with coded outlier values

Variable name	Data file	Variable description
R04_AM0065	Adult	Number of years lived in the United States
R04_AC1009_NN	Adult	How long since last smoked a cigarette - Number
R04_AV1009_NN	Adult	How long since you last took a puff from an electronic nicotine product - Number
R04_AG1009TC_NN	Adult	How long since last smoked a traditional cigar - Number
R04_AG1009CG_NN	Adult	How long since last smoked a cigarillo - Number
R04_AG1009FC_NN	Adult	How long since last smoked a filtered cigar - Number
R04_AG1009TJ_NN	Adult	How long since last smoked a traditional cigar as a blunt - Number
R04_AG1009GJ_NN	Adult	How long since last smoked a cigarillo as a blunt - Number
R04_AG1009FJ_NN	Adult	How long since last smoked a filtered cigar as a blunt - Number
R04_AP1009_NN	Adult	How long since you last smoked a pipe filled with tobacco - Number
R04_AH1009_NN	Adult	How long since you last smoked tobacco in a hookah - Number
R04_AU1009_NN	Adult	How long since you last used snus - Number
R04_AS1009_NN	Adult	How long since you last used smokeless tobacco - Number
R04_AC1006_RS	Adult	Age when first smoked part or all of a cigarette
R04_AC1007_NB	Adult	Age when first started smoking cigarettes fairly regularly
R04_AC9002_NN	Adult	How long smoking/smoked cigarettes fairly regularly - Number
R04_AC9029_NN	Adult	How long since you last smoked a cigarette - Number (Some day or current experimental smokers)
R04_AC1024_NN	Adult	Time to first cigarette after waking on days smoked - Number (current smokers)
R04_AC1024_30D_NN	Adult	Time to first cigarette after waking on days smoked - Number (non-current smokers)
R04_AC1051MC_NN	Adult	How long smoked regular brand of manufactured cigarettes - Number
R04_AV1006_RS	Adult	Age when first used an electronic nicotine product, even one or two times
R04_AV1007_NB	Adult	Age when first started using electronic nicotine products fairly regularly
R04_AV1020_NB	Adult	Age when first started using electronic nicotine products every day
R04_AV9045	Adult	Number of milliliters of e-liquid your tank system holds/held
R04_AV9029_NN	Adult	Time since you took last puff from an electronic nicotine product - Number (Some day or current experimental users)
R04_AV1026	Adult	Average number of times you pick up your electronic nicotine product to take one or more puffs each day on days used
R04_AV1027	Adult	Number of puffs you take each time you pick up your electronic nicotine product to use it
R04_AV1024_NN	Adult	Time to first electronic nicotine product puff after waking - Number (current users)
R04_AV1024_30D_NN	Adult	Time to first electronic nicotine product puff after waking - Number (non-current users)
R04_AV1138_D	Adult	How much usually pay/paid for a box or pack of [electronic nicotine products/electronic nicotine cartridges/milliliters of e-liquid] - Dollars
R04_AV1138_C	Adult	How much usually pay/paid for a box or pack of [electronic nicotine products/electronic nicotine cartridges/milliliters of e-liquid] - Cents

Table C-4. Wave 4 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R04_AV1051_NN	Adult	How long used regular brand of [electronic nicotine products/electronic nicotine cartridges/e-liquid] - Number
R04_AM0103	Adult	Number of puffs of marijuana taken from an electronic nicotine product today/yesterday/day before yesterday
R04_AJ9029TC_NN	Adult	Time since last smoked traditional cigar as blunt - Number (Some day or current experimental smokers)
R04_AJ1023TC	Adult	In past 30 days, average number of traditional cigars as blunts smoked per day on days smoked
R04_AJ9029CG_NN	Adult	Time since last smoked cigarillo as blunt - Number (Some day or current experimental smokers)
R04_AJ1023CG	Adult	In past 30 days, average number of cigarillos as blunts smoked per day on days smoked
R04_AJ9029FC_NN	Adult	Time since last smoked filtered cigar as blunt - Number (Some day or current experimental smokers)
R04_AG1006TC_RS	Adult	Age when first smoked part or all of a traditional cigar, even one or two puffs
R04_AG1007TC_NB	Adult	Age when first started smoking traditional cigars fairly regularly
R04_AG9029TC_NN	Adult	Time since last smoked traditional cigars - Number (Some day or current experimental smokers)
R04_AG1024TC_NN	Adult	Time to first traditional cigar after waking on days smoked - Number (current smokers)
R04_AG1024TC_30D_NN	Adult	Time to first traditional cigar after waking on days smoked - Number (non-current smokers)
R04_AG1006CG_RS	Adult	Age when first smoked part or all of a cigarillo, even one or two puffs
R04_AG1021CG	Adult	Average number of cigarillos now smoked each day
R04_AG9029CG_NN	Adult	Time since last smoked cigarillos - Number (Some day or current experimental smokers)
R04_AG1024CG_NN	Adult	Time to first cigarillo after waking on days smoked - Number (current smokers)
R04_AG1024CG_30D_NN	Adult	Time to first cigarillo after waking on days smoked - Number (non-current smokers)
R04_AG1040CG	Adult	Number of cigarillos that come/came in box or pack usually buy/bought
R04_AG1006FC_RS	Adult	Age when first smoked part or all of a filtered cigar, even one or two puffs
R04_AG1007FC_NB	Adult	Age when first started smoking filtered cigars fairly regularly
R04_AG9029FC_NN	Adult	Time since last smoked filtered cigars - Number (Some day or current experimental smokers)
R04_AG1024FC_NN	Adult	Time to first filtered cigar after waking on days smoked - Number (current smokers)
R04_AG1024FC_30D_NN	Adult	Time to first filtered cigar after waking on days smoked - Number (non-current smokers)
R04_AP1006_RS	Adult	Age when you first smoked a pipe filled with tobacco, even one or two puffs
R04_AP1007_NB	Adult	Age when first started smoking a pipe filled with tobacco fairly regularly
R04_AP9029_NN	Adult	Time since last smoked a pipe filled with tobacco - Number (Some day or current experimental smokers)

Table C-4. Wave 4 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R04_AP1024_NN	Adult	Time to first pipe bowl after waking on days smoked - Number (current smokers)
R04_AP1024_30D_NN	Adult	Time to first pipe bowl after waking on days smoked - Number (non-current smokers)
R04_AP1051_NN	Adult	How long smoked regular brand of pipe tobacco - Number
R04_AH1006_RS	Adult	Age when first smoked tobacco in a hookah, even one or two puffs
R04_AU1006_RS	Adult	Age when first used snus, even one or two times
R04_AU9029_NN	Adult	How long since you last used snus - Number (Some day or current experimental users)
R04_AU1024_NN	Adult	Time to first snus product after waking on days used - Number (current users)
R04_AU1024_30D_NN	Adult	Time to first snus product of the day after waking on the days used - Number (non-current users)
R04_AS1006_RS	Adult	Age when you first used smokeless tobacco, even one or two times
R04_AS1021	Adult	Average number of times smokeless tobacco now used each day
R04_AS1022	Adult	In past 30 days, number of days used smokeless tobacco
R04_AS9029_NN	Adult	Time since last used smokeless tobacco - Number (Some day or current experimental users)
R04_AS1024_NN	Adult	Time to first use of smokeless tobacco after waking on days used - Number (Current users)
R04_AS1051_NN	Adult	How long used regular brand of smokeless tobacco - Number
R04_AY0010_NN	Adult	Time to first use of any tobacco product after waking - Number
R04_AN0115E	Adult	In past 12 months, number of times tried to quit using electronic nicotine products
R04_AN0130E_NN	Adult	In past 12 months, length of time you stopped using electronic nicotine products because you were trying to quit - Number
R04_AN0135E_VALUE_INFO	Adult	Type of response provided for R04_AN0135E
R04_AN0145E_NN	Adult	In past 12 months, longest time period for which you stopped using electronic nicotine products because you were trying to quit - Number
R04_AN0175E_NN	Adult	How long ago did you stop using NRT - Number (current established, recent former established or current experimental electronic nicotine product users)
R04_AN0115	Adult	In past 12 months, number of times tried to quit smoking/using tobacco product(s)
R04_AN0130_NN	Adult	In past 12 months, length of time you stopped smoking/using tobacco product(s) because you were trying to quit - Number
R04_AN0135_VALUE_INFO	Adult	Type of response provided for R04_AN0135
R04_AN0145_NN	Adult	In past 12 months, longest time period for which you stopped smoking/using tobacco product(s) because you were trying to quit - Number
R04_AN0251_NN	Adult	How long been using NRT - Number (past 12 month non-electronic tobacco quitters or quit attempters)
R04_AN0175_NN	Adult	How long ago did you stop using NRT - Number (current established or recent former established non-electronic tobacco users)
R04_AN0195_NN	Adult	Length of time used prescription medication during last tobacco product(s) quit attempt - Number

Table C-4. Wave 4 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R04_AX0068	Adult	In past 7 days, number of hours that you were in close contact with others when they were smoking
R04_AX0679_FT	Adult	Height without shoes: Feet
R04_AX0679_IN	Adult	Height without shoes: Inches
R04_AX0316	Adult	Height without shoes: Meters
R04_AX0109	Adult	Current weight: Pounds
R04_AX0312	Adult	Current weight: Kilograms
R04_AX0242_NN	Adult	On days when you do physical activity or exercise of at least moderate intensity, how long do you do these activities - Number
R04_AX0114_NB	Adult	Age when you were first told you had high blood pressure
R04_AX0116_NB	Adult	Age when you were first told you had congestive heart failure
R04_AX0117_NB	Adult	Age when you were first told you had a stroke
R04_AX0112_NB	Adult	Age when you were first told you had a heart attack or needed bypass surgery
R04_AX0120_NB	Adult	Age when you were first told you had COPD
R04_AX0121_NB	Adult	Age when you were first told you had chronic bronchitis
R04_AX0123_NB	Adult	Age when you were first told you had emphysema
R04_AX0124_NB	Adult	Age when you were first told you had asthma
R04_AX0695_NB	Adult	Number of permanent teeth removed because of tooth decay or gum disease
R04_AX0695_12M	Adult	In past 12 months, number of permanent teeth removed because of tooth decay or gum disease
R04_AX0725	Adult	Number of times you brush your teeth in one day
R04_AX0696	Adult	In the last seven days, number of times used dental floss or any other device to clean between teeth
R04_AX0729_NN	Adult	Time since you had your teeth cleaned by a dentist, hygienist, or other health professional - Number
R04_AX0131_NB	Adult	Age when you were first told you had gum disease
R04_AX0133_NB	Adult	Age when you were first told you had pre-cancerous oral lesions
R04_AX0280_NB	Adult	Age when you were first told you had diabetes, sugar diabetes, high blood sugar or borderline diabetes
R04_AX0143_NB	Adult	Age when you were first told you had an ulcer
R04_AX0150_NB	Adult	Age when you were first told you had osteoporosis
R04_AX0703	Adult	Age when first told you had macular degeneration
R04_AX0757	Adult	Age when you were first told by a doctor, therapist or other mental health professional that you had schizophrenia, schizoaffective disorder or psychosis
R04_AX0762	Adult	Age when you were first told by a doctor, therapist or other mental health professional that you had a psychotic illness or episode
R04_AX0137_NN	Adult	Weeks/Months pregnant - Number
R04_AX0086_RS	Adult	Age when first drank alcohol at all, counting small tastes or sips
R04_AX0074_RS	Adult	Age when first alcoholic drink was consumed, other than small tastes or sips
R04_AX0079_RS	Adult	Age when first used marijuana, hash, THC, grass, pot or weed
R04_PM0058_NEW_AGE	Youth	Age of other parental figure/guardian
R04_PM0061_NEW_AGE	Youth	Age of second other parental figure/guardian
R04_PM0065_NN	Youth	Number of years lived in the United States (parent respondent)
R04_YM0065	Youth	Number of years lived in the United States (youth)
R04_YC1006_NB	Youth	Age when first tried cigarette smoking, even one or two puffs

Table C-4. Wave 4 questionnaire variables with coded outlier values (continued)

Variable name	Data file	Variable description
R04_YV1006_NB	Youth	Age when you first tried an electronic nicotine product, even one or two times
R04_YV1007	Youth	Age when first started using electronic nicotine products fairly regularly
R04_YV0103	Youth	Number of puffs taken from an electronic nicotine product today/yesterday/day before yesterday
R04_YV9040	Youth	In past 30 days, number of [electronic nicotine products/nicotine cartridges/milliliters of e-liquid] used per day on days used an electronic nicotine product
R04_YV1026	Youth	Average number of times you pick up your electronic nicotine product to use it for one or more puffs on days that you use
R04_YG1006TC_NB	Youth	Age when first tried a traditional cigar, even one or two puffs
R04_YG1007TC	Youth	Age when first started smoking traditional cigars fairly regularly
R04_YG1006FC_NB	Youth	Age when first tried a filtered cigar, even one or two puffs
R04_YH9006	Youth	Average number of times you smoke tobacco in a hookah each day
R04_YU1006_NB	Youth	Age when you first tried snus, even one or two times
R04_YS0103	Youth	Number of times used smokeless tobacco today/yesterday/day before yesterday
R04_YN0329	Youth	Number of Chantix, varenicline, Wellbutrin, Zyban or Bupropion taken today/yesterday/the day before yesterday (past 12 month tobacco users)
R04_YX0068	Youth	In past seven days, number of hours that you were in close contact with others when they were smoking
R04_YT0007_FT	Youth	Current height: Feet
R04_YT0007_IN	Youth	Current height: Inches
R04_YX0311	Youth	Current height: Meters
R04_YT0008	Youth	Current weight: Pounds
R04_YX0312	Youth	Current weight: Kilograms
R04_YN0309H	Youth	Number of nicotine patches, gum, inhaler, nasal spray, lozenges or pills used today/yesterday/the day before yesterday (past 12 month tobacco non-users)
R04_YN0329H	Youth	Number of Chantix, varenicline, Wellbutrin, Zyban or bupropion taken today/yesterday/the day before yesterday (past 12 month tobacco non-users)
R04_YX0086_NB	Youth	Age when first drank alcohol at all
R04_YX0074_NB	Youth	Age when consumed first alcoholic drink
R04_YX0079_NB	Youth	Age when first used marijuana, hash, THC or grass