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

Data and Biospecimen Collection Nonresponse Bias Analysis for Wave 4

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 #: HHSN271201600001C

PATH Study Wave 4 Data and Biospecimen Collection

Nonresponse Bias Analysis Report

November 28, 2018
Updated September 9, 2019

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

Table of Contents

Appendix **Page**

Table

Table **Page**

Figure

1. Introduction

Westat prepared this report as a reference document for researchers using the Population Assessment of Tobacco and Health (PATH) Study interview and biomarker data. This report focuses on the Wave 4 response rates and potential nonresponse biases for select demographic and outcome measures. Similar reports are also available for Wave 1 through Wave 3, and can be found at the website for the National Addiction & HIV Data Archive Program (NAHDAP).

In this document, the term "participants" is used to indicate people who agreed to be a 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.).

At Wave 4, the PATH Study data support analyses of two cohorts, referred to as the "Wave 1 Cohort" and "Wave 4 Cohort." The Wave 1 Cohort includes all Wave 1 respondents. These study participants were selected to represent adults ages 18 and older and youth ages 9-17 in the U.S. civilian, noninstitutionalized population (CNP) at the time of Wave 1. The Wave 1 Cohort is to be used for longitudinal analysis at Wave 4. To maximize sample size for such analyses, all Wave 1 respondents were eligible for Wave 4 data collection unless they were incarcerated, deceased, or resided outside of the U.S. at the time of Wave 4.

The newly established Wave 4 Cohort consists of two components. The first component is the Wave 4 adult and youth respondents from the Wave 1 Cohort who were in the U.S. CNP at the time of Wave 4. The second component is a replenishment sample designed to supplement the Wave 1 sample. The Wave 4 Cohort serves the purpose of cross-sectional analysis at Wave 4.

This report summarizes the Wave 4 nonresponse bias analysis conducted separately for the two cohorts. The analysis for the Wave 1 Cohort is similar to that conducted for Wave 3. The analysis for the Wave 4 Cohort is similar to that conducted for the Wave 1 Cohort at Wave 1. The potential for nonresponse bias in Wave 4 estimates based on the Wave 1 Cohort is evaluated conditioning on

Wave 1 response, so a brief summary of the findings from the Wave 1 nonresponse bias analyses is provided below:

- For the Wave 1 household screener and adult interview, most demographic and socio-economic characteristics of the respondents in Wave 1 aligned with estimates from the 2013 American Community Survey (ACS) when using the basic design weights, also known as the inverse-of-probability-of-selection (IPS) weights for the PATH Study. Exceptions were found for single-person households, sex, education, and ethnicity when comparing Wave 1 estimates using the IPS weights to the 2013 ACS estimates. Estimates of cigarette-smoking rates among adults in Wave 1 were within the range of estimates found in other national health studies. When the estimates were adjusted for nonresponse using the Wave 1 final weights, they more closely approximated the ACS estimates and adult cigarette-smoking rates remained essentially the same.

- For the Wave 1 youth interview, most demographic characteristics of respondents were consistent with the estimates from the 2013 ACS, with the exception of ethnicity, when using the IPS weights. When the estimates were adjusted for nonresponse among youth, they more closely approximated the 2013 ACS estimates, but the ever-tried cigarette-smoking rates for all youth in Wave 1 remained 2 to 12 percentage points lower than those found in other national studies.

- For the Wave 1 urine and blood collections, most of the demographic and socio-economic characteristics of specimen providers generally aligned with estimates of these characteristics from the 2013 ACS, when using the IPS weights. In addition, when the estimates were adjusted for nonresponse, they were found to approximate the ACS estimates more closely.

This report is organized as follows: Chapter 2 provides an overview of the PATH Study sample design. Chapter 3 describes the methodology used for the analyses in this report. Chapters 4 and 5 present the results of nonresponse bias analyses for the Wave 1 Cohort and Wave 4 Cohort, respectively, with respect to the interview data and biospecimen collection data. Chapter 6 summarizes the findings and discusses their implications.

2. Overview of Sample Design and Data Collection

This chapter provides an overview of the sample design for the PATH Study. The PATH Study is a nationally representative, longitudinal cohort study of tobacco-use behaviors and related health outcomes among adults and youth in the United States. Interviews are conducted with respondents ages 12 years and older in annual or biennial waves. The study's design allows for the longitudinal assessment of within-person and between-person change in the use of tobacco products, including initiation, cessation, relapse, and transitions between products, as well as in related health effects associated with use patterns.

At Wave 1, a four-stage, stratified probability sample design was used to select adults ages 18 and older and youth ages 12 to 17 from the U.S. CNP; an additional "shadow sample" of youth ages 9 to 11 was selected to be interviewed after they turn 12 years of age in later waves. All Wave 1 sample respondents together form the Wave 1 Cohort.

At Wave 4, the original set of Wave 1 sample respondents was replenished with a probability sample of adults, youth, and shadow youth ages 10 to 11[1] selected from the U.S. CNP at the time of Wave 4. The replenishment effort had two objectives: supplementation of the Wave 1 sample to address attrition and the changing composition of the study's adult respondents as Wave 1 youth (who were not oversampled with respect to tobacco use) reach age 18, and selection of a shadow sample for fielding in future waves. The replenishment effort also gave persons who had been overseas or in the military or an institutional setting at the time of Wave 1 (but were no longer at Wave 4) a chance of selection for the study. Because this "replenishment sample" was 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 sample who were in the CNP at the time of Wave 4. This combined set of Wave 4 respondents forms the Wave 4 Cohort.

[1] The difference in the shadow youth age range between the Wave 1 sample (ages 9 to 11) and the Wave 4 replenishment sample reflects the planned timings of (a) future data collections among the full PATH Study sample and (b) efforts to replenish the sample periodically. After Wave 4, data collection with the full sample is planned to occur biennially.

Figure 1 illustrates the components of the Wave 1 and Wave 4 Cohorts. Table 2-1 shows the sample sizes for the two cohorts.

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

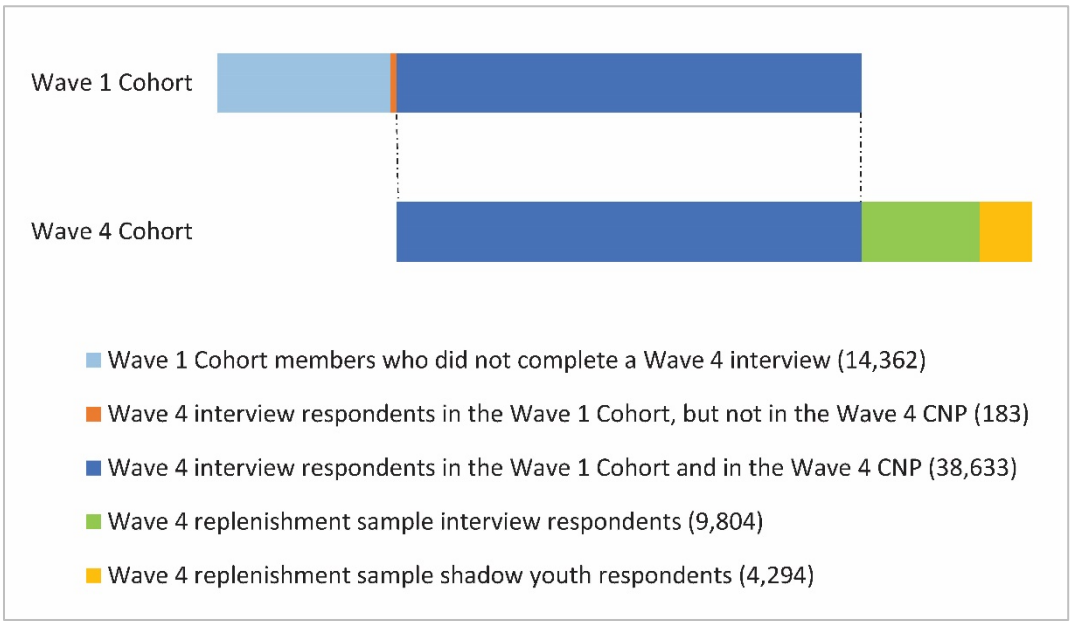

Table 2-1. Wave 1 Cohort and Wave 4 Cohort sample sizes

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

Key features of the Wave 1 and Wave 4 sample designs are summarized below. Further details can be found in Chapter 2 of the PATH Study Restricted Use Files User Guide (https://www.icpsr.umich.edu/icpsrweb/NAHDAP/studies/36231/datadocumentation).

2.1 Wave 1

The target population of the PATH Study at Wave 1 is the U.S. CNP nine years of age and older at the time of Wave 1 (2013–2014). Active duty military personnel and those residing in an institutional setting at the time of Wave 1 were excluded. College students living away from home during the school year were identified as members of their permanent residence (e.g., parents' home). For Wave 1, a four-stage stratified area probability sample design was used with a two-phase design for sampling adults at the final stage. The sampling rates for adults varied by age, race, and tobacco-use status. 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 (consisting of one or more census blocks) were formed and then a systematic sample of these segments was drawn. At the third stage, a sample of addresses within sampled segments was drawn from listings of addresses; the main source of these addresses was obtained from the U.S. Postal Service (USPS) Computerized Delivery Sequence Files (CDSFs). The CDSFs provide very high coverage of the residential addresses in the U.S. The fourth stage was the random selection of persons within sampled households.

For within-household selection, a roster of all the members in the sampled household was constructed using the household screener. An adult household member, the household screener respondent, was asked to list members of the household and provide demographic as well as, for each adult, tobacco-use information.[2] This information was used for sampling three main groups of interest:

- Adults ages 18 and older (up to two adults per household);
- Children ages 12 to 17 (i.e., "youth," generally up to two per household);[3] and
- Children ages 9 to 11 (i.e., "shadow youth," generally up to two per household) to be interviewed in later waves after reaching 12 years of age.

[2] The household screener collected information on adult household members' use of several different types of tobacco products. For example, it collected information on current use of products with relatively high prevalence or well-established use, such as cigarettes, cigars, and pipes; and on ever use of products with relatively low prevalence or newly emerging use, such as electronic cigarettes or e-cigarettes.

[3] Given a special analytic interest in multiple-birth youth, the shadow youth and youth sampling procedures were modified when households containing multiple-birth youths were encountered so that the multiple-birth youths 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.

Two-phase sampling was used for adult selection due to potential misreporting by the household screener respondent of the tobacco-use status of adult household members. Phase 1 sampling depended on the age, race, and tobacco-use information obtained from the household informant during the household screener. Phase 2 sampling was based on self-reported age, race, and tobacco-use status, obtained by interviewing the individuals sampled at Phase 1. Sampling rates for the two phases were designed to oversample young adults (ages 18 to 24) and adult tobacco users of all ages, thereby increasing precision for estimates relating to these subgroups.

PATH Study Wave 1 interview data and biospecimen collections began September 12, 2013 and ended December 14, 2014. For Wave 1, 32,320 adult interviews and 13,651 youth interviews were completed. All adult interview respondents were asked to provide urine and blood specimens; 21,801 provided a urine specimen and 14,520 provided a blood specimen. The Wave 1 Data and Biospecimen Collection Nonresponse Bias Analysis Report is available on the NAHDAP website.

2.2 Wave 2 and Wave 3

There was no additional sampling for Wave 2 or Wave 3 of the PATH Study. Wave 2 was the first follow-up wave of Wave 1 respondents. Wave 1 respondents were eligible for Wave 2 if they were still residents of the U.S. and not incarcerated.

During the Wave 2 data collection period, attempts were made to contact the Wave 1 adult and youth respondents as well as members of the shadow youth sample established at Wave 1. Shadow youth who turned age 12 by Wave 2 and were permitted by a parent or guardian to participate in the study were asked for assent to be interviewed for the first time at Wave 2. Similarly, persons in the youth sample at Wave 1 who reached age 18 by Wave 2 were asked to complete the adult interview and to provide urine and blood specimens.

The PATH Study Wave 2 interview data and biospecimen collections began October 23, 2014 and ended October 30, 2015. For Wave 2, 28,362 adult interviews and 12,172 youth interviews were completed. The study subsampled 14,465 adults for urine collection at Wave 2 from adults who completed an adult interview at Wave 1; among these adults, 12,561 completed the Wave 2 interview and 12,109 provided urine specimens again at Wave 2. The study also collected 1,587 urine specimens and 908 blood specimens from the 1,915 persons who had completed the PATH Study

adult interview for the first time (referred to as "first-time adult interview respondents") at Wave 2. The Wave 2 Data and Biospecimen Collection Nonresponse Bias Analysis Report is available on the NAHDAP website.

Wave 3 was the second follow-up wave of the PATH Study for the Wave 1 Cohort. Wave 1 respondents were eligible for Wave 3 if they were still residents of the U.S. and not incarcerated. Wave 2 nonrespondents were fielded for Wave 3 unless they were deceased, had moved out of the U.S. permanently, had specifically requested withdrawal from the study, were firm or hostile refusers at Wave 2, had a language barrier,[4] or had a physical or mental disability or chronic illness that prevented participation in the study. Shadow youth from previous waves who turned age 12 by Wave 3 and were permitted by a parent or guardian to participate in the study were asked for assent to be interviewed for the first time at Wave 3. Similarly, youth from previous waves who reached age 18 by Wave 3 were asked to complete the adult interview and to provide urine and blood specimens.

The PATH Study Wave 3 interview data and biospecimen collections began October 19, 2015 and ended October 23, 2016. For Wave 3, 28,148 adult interviews and 11,814 youth interviews were completed. The study subsampled 13,700 adult respondents for urine collection at Wave 3 from adults who had completed an adult interview at a prior wave; among these adults, 13,338 provided urine specimens again at Wave 3. The study also collected 1,641 urine specimens and 835 blood specimens from the 1,907 first-time adult interview respondents in Wave 3. The Wave 3 Data and Biospecimen Collection Nonresponse Bias Analysis Report is available on the NAHDAP website.

2.3 Wave 4

For members of the Wave 1 Cohort, Wave 4 of the PATH Study was the third follow-up wave. Wave 1 respondents were eligible for Wave 4 if they were still residents of the U.S. and not incarcerated. Nonrespondents from previous waves were fielded for Wave 4 unless they were deceased, had moved out of the U.S. permanently, had specifically requested withdrawal from the

[4] The PATH Study interviews can be conducted in English or Spanish. At Wave 2, the cases assigned a final nonresponse status due to language problems included two groups: (a) Wave 1 respondents who mentioned difficulty completing the Wave 1 interview in English or Spanish as their main reason for not completing an interview at Wave 2; and (b) Wave 1 shadow youth who did not feel comfortable completing an interview in English or Spanish when asked to do so by the study for the first time.

study, were firm or hostile refusers at a previous wave, had a language barrier, or had a physical or mental disability or chronic illness that prevented participation in the study. In addition, those who had refused participation at both Wave 2 and Wave 3 were not fielded for data collection.[5] (Note that refusal is a specific type of nonresponse, so some nonrespondents to Wave 2 and Wave 3 who did not refuse at both those waves were fielded for Wave 4.) Shadow youth from previous waves who turned age 12 by Wave 4 and were permitted by a parent or guardian to participate in the study were asked for assent to be interviewed for the first time at Wave 4. Similarly, youth from previous waves who reached age 18 by Wave 4 were asked to complete the adult interview and to provide urine and blood specimens. Unlike in previous waves, youth interview respondents were asked to provide a urine specimen.

At Wave 4, an additional sample was selected from the U.S. CNP to account for the aging of Wave 1 respondents and sample attrition among the Wave 1 Cohort. 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. A total of 174,273 mailing addresses were selected. To meet the needs for the Wave 4 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 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. However, the sampling rates were designed to bring the Wave 4 adult and youth sample sizes up to Wave 1 levels by sampling domain. This required consideration for the expected combined effect of the aging of Wave 1 respondents 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 given that 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. All adult interview respondents from the

[5] For youth, refusal could have come from the youth or their parent/guardian.

replenishment sample were asked to provide urine and blood specimens, and all youth interview respondents from this sample were asked to provide a urine specimen.

The PATH Study Wave 4 interview data and biospecimen collections began December 1, 2016 and ended January 3, 2018. A full summary of the results of these data collection efforts is provided in Table 2-2 and select counts are mentioned in the following text. Among the Wave 1 Cohort members, 27,757 adult interviews and 11,059 youth interviews were completed at Wave 4. Data were also collected from 6,065 adults and 3,739 youth in the replenishment sample, for a combined total of 33,822 Wave 4 adult interviews and 14,798 Wave 4 youth interviews. When restricted to respondents in the Wave 4 Cohort (i.e., excluding those not in the U.S. CNP at the time of Wave 4), 33,644 adult interviews and 14,793 youth interviews were completed.

The study subsampled 14,811 adult respondents for urine collection at Wave 4 from adults who had completed an adult interview at a prior wave; among these adults, 14,471 provided urine specimens. The study also collected 1,660 urine specimens and 890 blood specimens from the 1,900 members of the Wave 1 Cohort who were first-time adult interview respondents at Wave 4. Of the youth respondents from the Wave 1 Cohort, 9,892 provided a urine specimen. Among the replenishment sample, 4,915 of the adult interview respondents provided a urine specimen and 2,718 provided a blood specimen. Urine specimens were also collected from 3,205 of the replenishment sample youth interview respondents. Among the Wave 4 Cohort interview respondents (i.e., after combining the Wave 1 and Wave 4 replenishment samples and restricting to those in the U.S. CNP at the time of Wave 4), the study collected 20,928 urine and 3,602 blood specimens from adults and 13,095 urine specimens from youth.

Table 2-2. Wave 4 data collection summary

Cohort membership	Wave sampled for PATH Study	Group	Interviews conducted	Respondents asked to provide biospecimen(s)	Urine specimens collected	Blood specimens collected
Wave 1 Cohort	Wave 1	Adults who had completed an adult interview at a prior wave	25,857	14,811	14,471	N/A
		First-time adult interview respondents	1,900	1,900	1,660	890
		All adults	27,757	16,711	16,131	890
		Youth	11,059	11,059	9,892	N/A
Wave 4 Cohort	Wave 1	Adults who had completed an adult interview at a prior wave	25,697	14,706	14,367	N/A
		First-time adult interview respondents	1,882	1,882	1,646	884
	Wave 4	First-time adult interview respondents	6,065	6,065	4,915	2,718
	Wave 1 or Wave 4	All adults	33,644	22,653	20,928	3,602
	Wave 1	Youth	11,054	11,054	9,890	N/A
	Wave 4	Youth	3,739	3,739	3,205	N/A
	Wave 1 or Wave 4	All youth	14,793	14,793	13,095	N/A
Wave 1 Cohort and/or Wave 4 Cohort	Wave 1 or Wave 4	Adults	33,822	22,776	21,046	3,608
	Wave 1 or Wave 4	Youth	14,798	14,798	13,097	N/A

3. Methodology

This chapter first discusses the factors affecting nonresponse bias, and then describes the methods used for evaluating potential nonresponse bias in this report.

3.1 Factors Affecting Nonresponse Bias

Bias is the difference between a survey estimate and the actual population value. Although nonresponse bias can be a major concern in multi-stage household surveys, nonresponse does not necessarily lead to nonresponse bias in survey estimates. Assuming nonresponse to be a fixed property of an individual in the population, the nonresponse bias of an estimate can be expressed mathematically to show the relationship between the bias and two factors—the amount of nonresponse and the difference between respondents and nonrespondents (Groves, 2006):

$$Bias(\bar{y}_r) = (1 - r)(\bar{Y}_r - \bar{Y}_m)$$

where $\bar{y}_r$ is the estimated population mean based on the sample respondents only, r is the response rate in the target population, $\bar{Y}_r$ is the mean for respondents in the target population, and $\bar{Y}_m$ is the mean for nonrespondents in the target population.

That is, the magnitude of nonresponse bias depends on the correlation between response propensity and the measure of interest. Within the same survey, different estimates can be subject to different levels of nonresponse bias. Some measures, unrelated to the propensity to respond, can be immune from the biasing effect of nonresponse; others, in the same survey, can be subject to large biases (Groves, 2006). For example, if cigarette smokers were less inclined to respond to the PATH Study than nonsmokers, then the estimated prevalence of cigarette smoking would be subject to nonresponse bias, prior to any weighting adjustments for nonresponse. However, if cigarette smoking were unrelated to, say, number of years lived in the U.S., then the estimated average years of residency may be unaffected by the different response propensities of smokers and nonsmokers. If propensity to respond is completely random, then nonresponse reduces sample sizes for analysis but does not lead to bias, even if the overall response rate is lower than desired.

In practice, survey practitioners often attempt to decrease potential nonresponse bias by not only increasing the overall response rate but also improving cooperation from "difficult to reach" subgroups. In addition, effective statistical adjustments can also help reduce potential nonresponse biases in some survey estimates.

3.2 Analyses for Evaluating Wave 4 Potential Nonresponse Bias

It is not always possible to measure the actual bias due to nonresponse; however, there are various approaches that can be used to help identify potential sources of nonresponse bias. At Wave 4, the Wave 1 Cohort data serve the purpose of longitudinal analysis, while the Wave 4 Cohort data support cross-sectional analysis. Therefore, separate sets of analyses were conducted for the two cohorts to assess potential nonresponse bias in estimates using data from Wave 4 of the PATH Study.

3.2.1 Comparison of Response Rates Across Subgroups

Although a response rate does not yield direct estimates of potential nonresponse biases on key measures, examining response rates by subgroups (for example, males versus females) may reveal sources of potential nonresponse bias. Wave 4 response rates were calculated for the Wave 1 Cohort and the Wave 4 replenishment sample separately, as described below. 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).

All Wave 1 respondents were eligible for Wave 4 unless they were incarcerated, deceased, or resided outside of the U.S. at the time of Wave 4 data collection. The Wave 4 interview response rates for the Wave 1 Cohort are conditioning on Wave 1 response. This should be kept in mind in the sections that follow (i.e., Sections 4.1.1 and 4.2.1). As in previous waves, persons 18 and older were asked to complete an adult interview and persons 12 to 17 were asked to complete a youth interview. A Wave 4 nonrespondent does not have a Wave 4 interview date, so his/her Wave 4 age and interview type (i.e., adult interview versus youth interview) were determined using the best available date of birth or age information. The latest such information was assumed the best for classifying age and interview type. Each Wave 1 Cohort member had an "anniversary month" for Wave 4 based on the interview date(s) of the study member(s) in his/her household at Wave 3 (or

Wave 2 or Wave 1). The age classification date for a Wave 4 nonrespondent was 1 month after the last day of his/her anniversary month or the final date of the Wave 4 data collection, whichever was earlier. A nonrespondent was considered eligible for a Wave 4 youth interview if his/her age was determined to be between 12 and 17 on the age classification date, or eligible for a Wave 4 adult interview if his/her age was determined to be 18 or older on the age classification date.

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. As demonstrated by results presented in Sections 4.1 and 4.2 for the Wave 1 Cohort, differences in response status prior to Wave 4 led to large differences in Wave 4 response rates. Furthermore, follow-up wave response rates are conditioning on Wave 1 response for those sampled at Wave 1, whereas response rates for those from the replenishment sample 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[6] and separate response rates were calculated for adults and youth from the Wave 4 replenishment sample. The household screener response rates in the main body of this report were computed for the AYS replenishment sample (from which all Wave 4 replenishment sample adults and youth were selected) at Wave 4.

In addition to the overall household screener response rate for the AYS replenishment sample, response rates are reported for subgroups defined by geography or neighborhood characteristics. For the Wave 4 replenishment sample, adult and youth interview response rates are reported for subgroups based on information collected from the household screener (e.g., age of sampled person). For the Wave 1 Cohort, the adult and youth interview response rates are conditioning on Wave 1 response, so the subgroups are based on Wave 1 information, with the exception of a subgroup defined by the study member's response statuses at Waves 2 and 3.

[6] 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.

Both weighted and unweighted response rates are reported for the household screener and adult and youth interviews. The unweighted response rate measures the success of field operations in obtaining responses from the sampled persons. The weighted response rate estimates the proportion of the population represented by the sampled persons that would have responded if they all had been asked to participate in the study, and thus provides a measure of the potential impact of nonresponse on the study estimates.[7] The basic design weights associated with sample selection, also known as the Wave 1 inverse-of-probability-of-selection (IPS) weights (for the Wave 1 Cohort) and the Wave 4 IPS weights (for the Wave 4 replenishment sample), were used for calculating the weighted response rates.

3.2.2 Wave 1 Cohort: Comparison to "Frame" Data from Wave 1

In a longitudinal study, the information collected in the baseline wave can be viewed as "frame" information for future waves, and thus used for evaluating potential nonresponse bias (Bose and West, 2002; Javitz and Wagner, 2005; Brownstein et al., 2009). For the PATH Study Wave 1 Cohort, estimates based on Wave 4 interview respondents were compared to the "frame" information from the baseline wave (Wave 1) for Wave 4 adults and youth, separately. This comparison helps identify characteristics that might be associated with nonresponse bias due to attrition between Wave 1 and Wave 4, after compensating for Wave 1 nonresponse and possible undercoverage.

The "frame" information included Wave 1 demographic and socio-economic characteristics, as well as Wave 1 tobacco-use status. The estimates were based on all the Wave 1 respondents who were eligible for the Wave 4 interview and were computed using the Wave 1 final weights (which were designed to reduce potential nonresponse bias from Wave 1).

Two sets of estimates were obtained from Wave 4 interview respondents. The first set of estimates was based on the Wave 1 final weights. Comparing this set of estimates to the "frame" data shows the representativeness of the Wave 4 responding sample before any statistical adjustment for attrition between Waves 1 and 4. The second set of estimates was based on the Wave 4 final weights that apply to all Wave 4 respondents in the Wave 1 Cohort regardless of their Wave 2 or Wave 3 response status (referred to as "single-wave weights"). The Wave 4 single-wave weights were

[7] Although the Wave 4 replenishment sample was not designed to support study estimates on its own, it is a nationally representative, probability-based sample, so the weighted response rates have a clear interpretation.

designed to reduce potential nonresponse bias due to attrition between Waves 1 and 4, so the second set of estimates shows the extent to which the statistical weighting adjustment might have helped reduce potential nonresponse bias at Wave 4. For reference, the terms "before Wave 4 weighting adjustment" and "after Wave 4 weighting adjustment" are used to refer to the two sets of estimates and comparisons in Sections 4.1 and 4.2. To facilitate interpretation of these two sets of estimates and comparisons, summary descriptions of the PATH Study statistical weighting adjustments for the Wave 1 Cohort are provided in the next two paragraphs.

Weighting adjustment is often used to account for differential response propensities across population subgroups. Among numerous sources, the Handbook on Household Surveys by the United Nations (2005, Chapter 6) and Särndal and Lundström (2005) discuss the methods and theory of using weight adjustments for nonresponse. For Wave 1, these adjustments were conducted at the household level and at the person level. The Wave 1 household-level weighting adjustments calibrated the estimates to household-level population estimates for census region and household composition and size from the 2013 ACS. Such weighting adjustments also correct for disparities among other characteristics that might be associated with the variables involved in the weighting adjustments. After accounting for household-level nonresponse, households with at least one person sampled for the PATH Study were identified, and each sampled person within a household was assigned the corresponding household weight with an adjustment to account for his/her within-household probability of selection. These weights were then adjusted to account for nonresponse to the Wave 1 adult or youth interview or non-participation in the shadow youth recruitment. After this adjustment for nonresponse, the weights were calibrated using a raking process to person-level population estimates from the 2013 ACS. Outlier values of the sample weights were trimmed, and the weights were re-raked after any such trimming. More details about the PATH Study Wave 1 weight construction can be found in Section 5.1.1 of the PATH Study RUF User Guide.

For Wave 4, approaches for creating the adult interview and youth interview single-wave weights were similar except that some auxiliary variables for nonresponse adjustment and raking differed by interview type. The final person-level weights assigned to Wave 1 respondents served as the initial weights for developing the single-wave weights for Wave 4 interview respondents. These weights were adjusted to account for nonresponse to the Wave 4 interview and the resulting weights were raked to control totals. Some of the control totals came from the 2013 ACS; others involving tobacco use were sample-based rather than population-based and reflected estimated Wave 1

population characteristics. Raking to sample-based control totals, often employed in longitudinal studies (see, for example, Brick, Lê, and West (2003)), can limit drifting in some important baseline characteristics that might arise through the applications of nonresponse adjustments over time. Lundström and Särndal (1999) provide theoretical discussions about sample-based calibration together with empirical evidence that such calibration can help reduce both variance and nonresponse bias. General discussions about the calibration method can be found in Särndal and Lundström (2005) and Särndal (2007). More details about the PATH Study Wave 4 single-wave weight construction for the Wave 1 Cohort appear in Section 5.1.4 of the PATH Study RUF User Guide.

3.2.3 Wave 4 Cohort: Comparison to External Data Sources

For the first three waves of the PATH Study, the Wave 1 Cohort is to be used to produce cross-sectional estimates (which are approximate in nature for Waves 2 and 3). For Wave 4, the Wave 4 Cohort serves the purpose of cross-sectional analysis. A widely used approach for assessing potential nonresponse bias in cross-sectional estimates is benchmarking to external sources. A strength of this method is that estimates from external sources are independent of the PATH Study. On the other hand, key outcomes from the PATH Study are not necessarily available from external sources; the measurements may not be exactly the same across studies, and the external sources may be subject to coverage and nonresponse errors as well (Groves, 2006).

Two types of estimates were obtained using the Wave 4 Cohort adult interview and youth interview data. First, weighted distributions of demographic and socio-economic variables were compared to those from the 2016 ACS. Second, several PATH Study cigarette-smoking measures were compared to those from other national studies of tobacco and health. The PATH Study measures a range of tobacco-use behaviors; many of these variables are not available in other studies. However, responses to the PATH Study questions on current cigarette smoking can be compared with estimates from other national studies that ask about cigarette-smoking behavior. Possible differences were examined between the cigarette-smoking estimates from the PATH Study and independent estimates of those quantities from other studies (for the most similar timeframes for which data were available).

The external data sources for cigarette smoking included the Tobacco Use Supplement to the Current Population Survey 2014-2015 (TUS-CPS 2014-2015), the National Health Interview Survey 2017 (NHIS 2017), the National Health and Nutrition Examination Survey 2015-2016 (NHANES 2015-2016), the National Survey on Drug Use and Health 2016 (NSDUH 2016), and the National Youth Tobacco Survey 2017 (NYTS 2017). Appendix A describes the questions used to define current cigarette smoking on each of these surveys as well as the PATH Study, and outlines differences in target populations among these surveys and the PATH Study.

For both the PATH Study and the external data sources for cigarette smoking, item nonresponse was handled by excluding respondents with missing values for an item from the counts and estimates regarding that item. The proportions of item missing values were generally very low (below 1 percent for adult estimates at the overall level and mostly below 2.5 percent for youth estimates at the overall level) in both the PATH Study and the surveys that were used for comparison purposes.[8]

As discussed in Chapter 2, the Wave 4 Cohort consists of the Wave 4 adult and youth respondents from the Wave 1 sample (with the exception of those not in the U.S. CNP at Wave 4) and the respondents from the replenishment sample. The Wave 1 sample and the Wave 4 replenishment sample were selected at different times from two different frames, so the probabilities of selection (and IPS weights) were computed for the two samples separately. However, the information needed to create the IPS weight for the Wave 4 Cohort is not available. This is because an adjustment to the IPS weights computed for the separate samples is needed for persons with multiple chances of selection for the Wave 4 Cohort. The information needed to make this adjustment (e.g., Wave 1 age, race, and tobacco-use status from which the probability of an adult being selected in Wave 1 could be determined) is known only for some replenishment sample respondents and is unknown for persons from the replenishment sample who did not respond at Wave 4. Therefore, Wave 4 Cohort estimates for the interview nonresponse bias analyses were generated using only the Wave 4 Cohort final weights (i.e., the Wave 4 cross-sectional weights).

A sequence of weighting steps involving the Wave 1 Cohort weights and Wave 4 replenishment sample weights was implemented to create the Wave 4 Cohort final weights, including nonresponse

[8] The proportion of item missing values for the NHANES youth data was slightly higher at 5.9 percent. However, note that the NSDUH 2016 public use files included only "usable" cases that met specified minimum item response requirements.

adjustments, preliminary raking, compositing, and final raking and trimming. The control totals were created using the 2016 ACS Public Use Microdata Sample (PUMS). For adults, 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. (Missing values in the variables used for raking were imputed.) Assuming that the variables used in weighting are correlated with the PATH Study's key outcome measures, the weighting adjustments serve the purpose of reducing potential nonresponse bias. More details about the creation of the Wave 4 Cohort cross-sectional weights can be found in Section 5.1.4 of the PATH Study RUF User Guide.

3.2.4 Analyses About Biospecimen Collections

Biospecimens provide a basis for the assessment of between-person differences and within-person changes over time in markers of tobacco exposure. They also allow for the detection of health conditions and disease processes potentially associated with the use of tobacco products. At Wave 4, all first-time adult interview respondents (including adult respondents of any age selected as part of the replenishment sample) were asked to provide urine and blood specimens. In addition, a subsample of other adult interview respondents in the Wave 1 Cohort were asked to provide a urine specimen. All youth interview respondents were asked to provide a urine specimen. Field interviewers collected the urine specimens; on separate visits, phlebotomists collected the blood specimens.

The nonresponse bias analyses for biospecimen collections were conducted separately for the Wave 1 Cohort and the Wave 4 Cohort using the same approach. Unweighted response rates were calculated, both overall and by subgroups, to assess the operational success of the urine and blood specimen collections among the designated Wave 4 adult interview or youth interview respondents. All Wave 4 biospecimen collection response rates presented in this report are conditioning on Wave 4 adult interview or youth interview completion.

To evaluate potential bias due to nonresponse to biospecimen collections, socio-demographic characteristics and Wave 4 tobacco-use status were compared between the first-time adult interview respondents and the biospecimen providers among them, separately for urine and blood. Similar comparisons were made between youth interview respondents and the urine providers among them.

For the Wave 1 Cohort, estimates were based on the Wave 4 single-wave (longitudinal) weights. For the Wave 4 Cohort, estimates were based on the Wave 4 cross-sectional weights. For both cohorts, the Wave 4 biospecimen collection nonresponse bias analyses are conditioning on Wave 4 adult interview or youth interview completion.

3.3 Estimation Method and Software Package

Fay's balanced repeated replication (BRR) method with a factor of 0.3 was used for variance estimation to account for the impact of the stratification and clustering involved in the PATH Study's sample design and the weighting adjustments made to the IPS weights. SAS software version 9.4 was used to calculate all the point estimates. Confidence intervals were estimated using the modified Wilson approach (Wilson, 1927; SAS Institute, 2013). The preferred approach for testing whether two point estimates differ is to examine the confidence interval for the difference between the two point estimates; if the 95 percent confidence interval does not include zero, it can be concluded that the difference between the two estimates is statistically significant at the 0.05 significance level. This is discussed in Heeringa, West, and Berglund (2010, Section 5.6.1). Another approach is to examine whether the confidence intervals for the two point estimates overlap; if the confidence intervals for two proportions do not overlap, then the difference between the two proportions is considered statistically significant. However, Schenker and Gentleman (2001) show that using the second approach results in a conservative test. For the analyses presented in this report, the first approach was used for all comparisons that involve only PATH Study estimates (i.e., all Wave 1 Cohort analyses and all biospecimen analyses), and the second, more conservative approach was used to compare estimates between the PATH Study and external sources (in Sections 5.2.2, 5.2.3, 5.3.2, and 5.3.3). No adjustments were made for multiple comparisons because all the statistical tests were based on pre-planned (i.e., not post-hoc) comparisons and are presented in this report.

4. Results of Wave 1 Cohort Nonresponse Bias Analyses at Wave 4

This chapter presents results of nonresponse bias analyses for the Wave 1 Cohort at Wave 4. The interview nonresponse bias analyses for the Wave 1 Cohort at Wave 4 are all conditioning on Wave 1 response. Sections 4.1.1 and 4.2.1 describe how the interview response rates vary across select subgroups. Sections 4.1.2 and 4.2.2 compare demographic and socio-economic characteristics as well as tobacco-use status between Wave 4 interview respondents and those eligible for the Wave 4 interview. Sections 4.3 and 4.4 focus on the analyses of biospecimen data collections. The results are presented separately for adults and youth.

4.1 Adult Interview

4.1.1 Response Rates Conditioning on Wave 1 Response

Among those in the Wave 1 Cohort, unweighted and weighted response rates were calculated for the Wave 4 adult interview using the following formulas:

$$RR_A = C_A/(C_A + N_A + e_A \times U_A)$$

$$e_A = (C_A + N_A)/(C_A + N_A + I_A)$$

where

RR_A = Wave 4 adult interview response rate;

C_A = number of Wave 4 adult interview complete cases;

N_A = number of Wave 4 adult interview nonrespondents known to be eligible;

U_A = number of Wave 4 adult interview nonrespondents with unknown eligibility status;

e_A = estimated proportion of eligible cases among the Wave 4 adult interview nonrespondents with unknown eligibility status; and

I_A = number of Wave 4 adult interview ineligible cases that were not permanently ineligible prior to Wave 4.

Unweighted counts and weighted counts based on the Wave 1 IPS weights were obtained for response status categories C_A, N_A, U_A, and I_A for unweighted response rates and weighted response rates, respectively. A small number of participants became permanently ineligible prior to Wave 4 due to permanent incarceration, death, or having permanently moved overseas; this set of cases, denoted as IP_A in Table 4-1, were not included in the equation for computing e_A.

Table 4-1 provides response rates for the Wave 4 adult interview. In addition to the overall row, response rates are shown by response status at earlier waves as well as by Wave 1 sex, age, race/ethnicity, and tobacco-use status. Persons with missing values for a particular characteristic were excluded from the response rate calculation for that characteristic.

The unweighted and weighted response rates for the Wave 4 adult interview were 72.9 percent and 73.5 percent, respectively. The unweighted and weighted response rates were also similar to each other for most subgroups. Table 4-1 shows moderately differential response rates across the various subgroups (e.g., male versus female) except for earlier waves' response status. First, the weighted response rate was 91.7 percent among those who responded in Waves 2 and 3. In contrast, the weighted response rate was only 15.7 percent among those who skipped one or both of these waves. Second, the weighted response rate was lower among males (71.2 percent) than among females (75.5 percent); this pattern is consistent with most household surveys (Groves and Couper, 1998; Stoop, 2005). Third, although it is generally harder to reach younger adults, the Wave 4 weighted response rates for Wave 1 age groups "under 18" and "18-24" (76.3 percent and 70.4 percent, respectively) were comparable to those for older age groups (73.4 percent for "25-44," 75.6 percent for "45-64," and 71.2 percent for "65 and above"). Fourth, non-Hispanic adults of other or multiple races (71.8 percent) and non-Hispanic White alone adults (73.0 percent) had moderately lower weighted response rates than Hispanic adults (74.9 percent) and non-Hispanic Black alone adults (76.4 percent). Finally, the weighted response rates were 3.1 percentage points lower for Wave 1 adult current established tobacco users than for Wave 1 adults who were not current established tobacco users, and 2.4 percentage points lower for Wave 1 youth who had ever used tobacco than for Wave 1 youth who were never users. Variation in response rates by subgroups is to be expected in large-scale data collection efforts. None of the differences among the demographic and tobacco-use subgroups in Table 4-1 causes serious concern about potential nonresponse bias in the Wave 4 adult interview estimates.

Table 4-1. Wave 4 adult interview response rates conditioning on Wave 1 response

Characteristic[a]	C_A: Completed (n)	IP_A: Permanently ineligible prior to Wave 4 (n)	I_A: Other ineligible (n)	N_A: Nonresponse known to be eligible (n)	U_A: Nonresponse with unknown eligibility status (n)	RR_A: Unweighted response rate (%)	RR_A: Weighted response rate (%)
Overall	27,757	400	519	4,133	6,292	72.9	73.5
Earlier waves' response status							
Responded in all earlier waves	26,072	0	268	1,824	869	90.7	91.7
Did not respond in at least one earlier wave	1,685	400	251	2,309	5,423	18.5	15.7
Wave 1 sex							
Male	13,509	229	328	2,247	3,451	70.6	71.2
Female	14,220	169	191	1,882	2,833	75.2	75.5
Wave 1 age group							
Under 18	5,131	7	38	737	868	76.2	76.3
18-24	6,238	13	129	1,017	1,713	69.8	70.4
25-44	8,042	56	113	1,157	1,900	72.6	73.4
45-64	6,362	153	128	874	1,301	74.7	75.6
65 and above	1,980	171	111	347	501	70.6	71.2
Wave 1 race/ethnicity							
Hispanic	5,477	34	90	673	1,143	75.3	74.9
Non-Hispanic White alone	15,694	266	301	2,658	3,710	71.3	73.0
Non-Hispanic Black alone	4,065	60	65	405	789	77.5	76.4
Non-Hispanic other race or multiple races	2,102	26	50	328	532	71.2	71.8
Wave 1 current established tobacco use[b] (Wave 1 adults only)							
Current established tobacco user	9,830	217	275	1,548	2,499	71.1	71.3
Not current established tobacco user	12,200	154	178	1,748	2,732	73.3	74.4
Wave 1 ever tobacco use[c] (Wave 1 youth only)							
Ever user	1,616	2	18	230	319	74.7	74.9
Never user	3,397	5	20	483	520	77.3	77.3

[a] For each Wave 1 characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] A tobacco user is defined as someone who uses one or more of the tobacco products covered by the Wave 1 adult interview. A 'current established user' of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly. The products covered by the Wave 1 adult interview are cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, e-cigarettes, and dissolvable tobacco.

[c] An 'ever user' is someone who has ever used one or more of the tobacco products covered by the Wave 1 youth interview. A 'never user' is someone who has never used any of those tobacco products. Ever use of a tobacco product is defined as having ever used the product, even one or two times. The products covered by the Wave 1 youth interview are cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, e-cigarettes, dissolvable tobacco, bidis, and kreteks.

4.1.2 Comparison Between Wave 4 Respondents and Those Eligible for Wave 4 Interview (Among the Wave 1 Cohort)

This section compares weighted estimates between the Wave 4 adult interview respondents in the Wave 1 Cohort and the Wave 1 respondents who were eligible for the Wave 4 adult interview. The weighted estimates cover demographic and socio-economic characteristics (shown in Table 4-2) as well as tobacco-use status (shown in Tables 4-3 and 4-4).

Section 3.2 describes the methods used for obtaining the results in Tables 4-2, 4-3, and 4-4. Each table shows two sets of comparisons between the Wave 4 adult interview respondents and those eligible for the Wave 4 adult interview. The first set is for the "before Wave 4 weighting adjustment" comparison, and the second set is for the "after Wave 4 weighting adjustment" comparison.

Table 4-2 compares Wave 1 demographic and socio-economic characteristics between the Wave 4 adult interview respondents in the Wave 1 Cohort and those Wave 1 respondents eligible for the Wave 4 adult interview. For each characteristic, the percentages sum to 100 percent over the categories associated with the characteristic. For example, for both the Wave 4 respondents and those eligible for the Wave 4 adult interview, the estimated percentage of males and the estimated percentage of females add to 100 percent. Although the confidence intervals for some estimated differences in the first set of comparisons in Table 4-2 did not include zero,[9] most of the differences between the Wave 4 respondents and those eligible for the Wave 4 adult interview were not large enough to be substantively meaningful. The most noticeable underrepresentation among the Wave 4 adult respondents was the male population, which tends to have lower response propensity than the female population in most household surveys (Groves and Couper, 1998; Stoop, 2005).

Table 4-2 has two columns labeled "Unweighted count." These columns represent the numerator for calculating an unweighted percentage for the category associated with the characteristic. Tables 4-3 and 4-4 each have two columns labeled "Sample size." These columns represent the denominator for calculating an unweighted estimate of current established tobacco use for the category associated with the characteristic. These definitions generalize to all tables in this report that use column headings of "Unweighted count" and "Sample size."

[9] Testing whether or not a 100(1-α) percent confidence interval for a difference between two population proportions includes zero is equivalent to a two-sided test of the null hypothesis that the difference is zero at the α significance level (see, for example, Hanushek and Jackson, 1977).

Table 4-2. Comparison of Wave 1 demographic and socio-economic characteristics between Wave 4 adult interview respondents and Wave 1 respondents who were eligible for Wave 4 adult interview

Wave 1 characteristic	Wave 1 respondents who were eligible for Wave 4 adult interview		Wave 1 respondents who completed Wave 4 adult interview				
				Before Wave 4 weighting adjustment		After Wave 4 weighting adjustment	
	Unweighted count	A: Weighted percentage, using Wave 1 final weights [95% confidence interval]	Unweighted count	B: Weighted percentage, using Wave 1 final weights [95% confidence interval]	Difference in weighted percentages [B – A] [95% confidence interval]	C: Weighted percentage, using Wave 4 final weights [95% confidence interval]	Difference in weighted percentages [C – A] [95% confidence interval]
Sex							
Male	19,207	48.1% [47.6%, 48.6%]	13,509	46.7% [46.1%, 47.2%]	-1.4% [-1.9%, -0.9%]	48.1% [47.5%, 48.7%]	0.0% [-0.1%, 0.0%]
Female	18,935	51.9% [51.4%, 52.4%]	14,220	53.3% [52.8%, 53.9%]	1.4% [0.9%, 1.9%]	51.9% [51.3%, 52.5%]	0.0% [-0.0%, 0.1%]
Age group							
Under 18	6,736	5.2% [4.9%, 5.4%]	5,131	5.4% [5.1%, 5.7%]	0.2% [0.1%, 0.3%]	5.1% [4.9%, 5.4%]	0.0% [-0.0%, -0.0%]
18-24	8,968	12.5% [12.2%, 12.8%]	6,238	12.0% [11.6%, 12.4%]	-0.5% [-0.7%, -0.3%]	12.5% [12.1%, 12.9%]	0.0% [-0.0%, 0.0%]
25-44	11,099	33.0% [32.6%, 33.5%]	8,042	32.9% [32.3%, 33.4%]	-0.2% [-0.6%, 0.3%]	33.1% [32.5%, 33.6%]	0.0% [0.0%, 0.1%]
45-64	8,537	32.9% [32.5%, 33.4%]	6,362	33.9% [33.4%, 34.5%]	1.0% [0.6%, 1.4%]	33.0% [32.5%, 33.6%]	0.1% [0.0%, 0.1%]
65 and above	2,828	16.4% [16.0%, 16.7%]	1,980	15.8% [15.4%, 16.3%]	-0.6% [-0.9%, -0.2%]	16.3% [15.9%, 16.7%]	-0.1% [-0.1%, -0.0%]
Race/ethnicity							
Hispanic	7,293	15.8% [15.4%, 16.1%]	5,477	16.0% [15.6%, 16.4%]	0.2% [-0.2%, 0.7%]	15.7% [15.3%, 16.2%]	0.0% [-0.1%, 0.0%]
Non-Hispanic White alone	22,062	65.3% [64.8%, 65.8%]	15,694	64.9% [64.3%, 65.4%]	-0.4% [-0.9%, 0.1%]	65.3% [64.7%, 65.8%]	0.0% [-0.1%, 0.1%]
Non-Hispanic Black alone	5,259	11.3% [11.0%, 11.7%]	4,065	11.7% [11.4%, 12.1%]	0.4% [0.1%, 0.7%]	11.3% [11.0%, 11.7%]	0.0% [-0.1%, 0.1%]
Non-Hispanic other race or multiple races	2,962	7.6% [7.4%, 7.9%]	2,102	7.4% [7.1%, 7.7%]	-0.2% [-0.5%, -0.0%]	7.7% [7.3%, 8.0%]	0.0% [-0.1%, 0.1%]

Table 4-2. Comparison of Wave 1 demographic and socio-economic characteristics between Wave 4 adult interview respondents and Wave 1 respondents who were eligible for Wave 4 adult interview (continued)

Wave 1 characteristic	Wave 1 respondents who were eligible for Wave 4 adult interview		Wave 1 respondents who completed Wave 4 adult interview				
				Before Wave 4 weighting adjustment		After Wave 4 weighting adjustment	
	Unweighted count	A: Weighted percentage, using Wave 1 final weights [95% confidence interval]	Unweighted count	B: Weighted percentage, using Wave 1 final weights [95% confidence interval]	Difference in weighted percentages [B – A] [95% confidence interval]	C: Weighted percentage, using Wave 4 final weights [95% confidence interval]	Difference in weighted percentages [C – A] [95% confidence interval]
Education							
Less than high school or GED	6,159	16.4% [16.0%, 16.8%]	4,504	16.2% [15.8%, 16.7%]	-0.2% [-0.5%, 0.2%]	16.4% [15.9%, 16.9%]	0.0% [-0.1%, 0.1%]
High school	7,345	24.2% [23.7%, 24.7%]	5,156	23.4% [22.8%, 24.0%]	-0.8% [-1.3%, -0.3%]	24.1% [23.6%, 24.7%]	-0.1% [-0.2%, -0.0%]
Some college, no degree	11,060	31.2% [30.7%, 31.7%]	7,979	31.3% [30.7%, 31.9%]	0.1% [-0.3%, 0.6%]	31.2% [30.6%, 31.8%]	0.0% [-0.1%, 0.1%]
Bachelor degree and above	6,683	28.2% [27.7%, 28.7%]	4,882	29.1% [28.5%, 29.6%]	0.9% [0.4%, 1.3%]	28.3% [27.7%, 28.9%]	0.1% [-0.0%, 0.2%]
Health insurance							
Yes	25,059	85.4% [84.8%, 85.9%]	18,137	85.6% [84.9%, 86.3%]	0.3% [-0.0%, 0.6%]	85.6% [84.9%, 86.2%]	0.2% [-0.1%, 0.5%]
No	5,998	14.6% [14.1%, 15.2%]	4,266	14.4% [13.7%, 15.1%]	-0.3% [-0.6%, 0.0%]	14.4% [13.8%, 15.1%]	-0.2% [-0.5%, 0.1%]

Table 4-3. Comparison of Wave 1 "current established tobacco-use" estimates between Wave 1 adults who completed Wave 4 adult interview and Wave 1 adults who were eligible for Wave 4 adult interview*

	Wave 1 adults who were eligible for Wave 4 adult interview		Wave 4 adult interview respondents who completed Wave 1 adult interview				
				Before Wave 4 weighting adjustment		After Wave 4 weighting adjustment	
Wave 1 characteristic[a]	Sample size	A: Weighted estimate, using Wave 1 final weights [95% confidence interval]	Sample size	B: Weighted estimate, using Wave 1 final weights [95% confidence interval]	Difference in weighted estimates [B – A] [95% confidence interval]	C: Weighted estimate, using Wave 4 final weights [95% confidence interval]	Difference in weighted estimates [C – A] [95% confidence interval]
Overall	30,557	23.7% [23.1%, 24.3%]	22,030	22.9% [22.2%, 23.6%]	-0.8% [-1.2%, -0.5%]	23.6% [23.0%, 24.3%]	-0.1% [-0.2%, -0.0%]
Sex							
Male	15,350	29.5% [28.7%, 30.4%]	10,645	28.0% [27.0%, 29.1%]	-1.5% [-2.0%, -1.0%]	29.3% [28.4%, 30.2%]	-0.2% [-0.3%, -0.1%]
Female	15,182	18.4% [17.7%, 19.1%]	11,367	18.4% [17.7%, 19.2%]	0.0% [-0.3%, 0.3%]	18.4% [17.7%, 19.1%]	0.0% [-0.1%, 0.1%]
Age group							
18-24	8,823	28.9% [27.5%, 30.3%]	6,148	27.2% [25.5%, 28.9%]	-1.7% [-2.4%, -1.0%]	28.7% [27.3%, 30.1%]	-0.2% [-0.2%, -0.1%]
25-44	10,875	28.9% [28.0%, 29.8%]	7,896	27.9% [26.7%, 29.1%]	-1.0% [-1.5%, -0.4%]	28.7% [27.7%, 29.8%]	-0.1% [-0.5%, 0.3%]
45-64	8,220	22.8% [21.9%, 23.8%]	6,134	22.1% [21.1%, 23.2%]	-0.7% [-1.2%, -0.3%]	22.7% [21.7%, 23.7%]	-0.2% [-0.5%, 0.2%]
65 and above	2,628	10.5% [9.3%, 11.7%]	1,848	10.0% [8.8%, 11.5%]	-0.4% [-1.0%, 0.2%]	10.5% [9.1%, 11.9%]	0.0% [-0.5%, 0.5%]
Race/ethnicity							
Hispanic	5,223	17.3% [16.3%, 18.4%]	3,880	17.4% [16.2%, 18.6%]	0.1% [-0.5%, 0.6%]	17.2% [16.1%, 18.5%]	-0.1% [-0.2%, 0.0%]
Non-Hispanic White alone	18,333	25.2% [24.3%, 26.1%]	12,986	23.9% [22.8%, 25.0%]	-1.3% [-1.7%, -1.0%]	25.1% [24.2%, 26.1%]	-0.1% [-0.2%, 0.0%]
Non-Hispanic Black alone	4,253	26.7% [25.4%, 28.1%]	3,252	27.8% [26.1%, 29.5%]	1.1% [0.2%, 2.0%]	26.7% [25.2%, 28.2%]	0.0% [-0.2%, 0.2%]
Non-Hispanic other race or multiple races	2,289	19.4% [17.8%, 21.0%]	1,602	17.5% [15.6%, 19.5%]	-1.9% [-3.1%, -0.6%]	18.8% [17.0%, 20.8%]	-0.5% [-1.0%, -0.1%]

*A tobacco user is defined as someone who uses one or more of the tobacco products covered by the Wave 1 adult interview. A 'current established user' of a given tobacco product is someone who currently uses the product every day or some days and for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly. The products covered by the Wave 1 adult interview are cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, e-cigarettes, and dissolvable tobacco.

[a] For each Wave 1 characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

Table 4-4. Comparison of Wave 1 "ever tobacco-use" estimates between Wave 1 youth who completed Wave 4 adult interview and Wave 1 youth who were eligible for Wave 4 adult interview*

Wave 1 characteristic[a]	Wave 1 youth who were eligible for Wave 4 adult interview		Wave 4 adult interview respondents who completed Wave 1 youth interview				
				Before Wave 4 weighting adjustment		After Wave 4 weighting adjustment	
	Sample size	A: Weighted estimate, using Wave 1 final weights [95% confidence interval]	Sample size	B: Weighted estimate, using Wave 1 final weights [95% confidence interval]	Difference in weighted estimates [B – A] [95% confidence interval]	C: Weighted estimate, using Wave 4 final weights [95% confidence interval]	Difference in weighted estimates [C – A] [95% confidence interval]
Overall	6,565	32.8% [31.3%, 34.3%]	5,013	32.1% [30.5%, 33.8%]	-0.7% [-1.3%, 0.0%]	32.9% [31.4%, 34.5%]	0.1% [-0.2%, 0.4%]
Sex							
Male	3,323	34.6% [32.7%, 36.5%]	2,522	33.8% [31.8%, 35.9%]	-0.7% [-1.7%, 0.3%]	34.3% [32.4%, 36.2%]	-0.3% [-0.7%, 0.2%]
Female	3,234	31.1% [29.3%, 33.0%]	2,484	30.5% [28.5%, 32.6%]	-0.6% [-1.5%, 0.4%]	31.5% [29.6%, 33.5%]	0.4% [-0.1%, 0.9%]
Race/ethnicity							
Hispanic	1,810	33.5% [30.5%, 36.6%]	1,425	32.3% [28.9%, 36.0%]	-1.2% [-2.4%, 0.1%]	33.5% [30.3%, 36.9%]	0.1% [-0.9%, 1.0%]
Non-Hispanic White alone	3,250	34.8% [32.7%, 37.0%]	2,392	34.1% [31.8%, 36.6%]	-0.7% [-1.7%, 0.4%]	34.8% [32.6%, 37.1%]	0.0% [-0.4%, 0.5%]
Non-Hispanic Black alone	851	27.5% [24.5%, 30.8%]	695	28.8% [25.5%, 32.3%]	1.2% [-0.3%, 2.8%]	29.0% [25.6%, 32.7%]	1.5% [0.3%, 2.7%]
Non-Hispanic other race or multiple races	587	27.3% [23.9%, 31.0%]	445	26.0% [22.1%, 30.2%]	-1.3% [-4.0%, 1.3%]	26.0% [22.2%, 30.3%]	-1.3% [-3.8%, 1.2%]

*An 'ever user' is someone who has ever used one or more of the tobacco products covered by the Wave 1 youth interview. A 'never user' is someone who has never used any of those tobacco products. Ever use of a tobacco product is defined as having ever used the product, even one or two times. The products covered by the Wave 1 youth interview are cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, e-cigarettes, dissolvable tobacco, bidis, and kreteks.

[a] For each Wave 1 characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

For the second set of comparisons in Table 4-2, the "after Wave 4 weighting adjustment" estimates based on the Wave 4 respondents aligned closely with the estimates based on those eligible for the Wave 4 adult interview. That is, the estimated differences in the last column of Table 4-2 were mostly zero (to one decimal place) and the confidence intervals all included or were in the proximity of zero. This is because during the Wave 4 raking process, the control totals accounted for estimated population characteristics based on the PATH Study Wave 1 interview data including age, sex, race, ethnicity, and education.

Tables 4-3 and 4-4 show the comparisons for the Wave 1 tobacco-use measures between Wave 4 adult interview respondents and those eligible for the Wave 4 adult interview, among the Wave 1 Cohort. The Wave 1 respondents who were eligible for the Wave 4 adult interview were divided into two groups based on their Wave 1 interview type (i.e., adult versus youth) and analyzed separately. Table 4-3 covers "current established tobacco use" for the Wave 1 adult interview respondents and Table 4-4 covers "ever tobacco use" for the Wave 1 youth interview respondents. Besides the overall estimates, these tables also show the estimated Wave 1 "current established use" and "ever use" rates for some demographic subgroups.

For the "before Wave 4 weighting adjustment" estimates, the differences between the Wave 4 respondents and those eligible for the Wave 4 adult interview were no more than 1.9 percentage points. In addition, most of the confidence intervals either included or were in the proximity of zero. For the "after Wave 4 weighting adjustment" estimates, the differences between the respondents and those eligible for the adult interview were mostly negligible. The largest difference was for the Wave 1 "ever tobacco use" measure for non-Hispanic Black alone adults who were youth in Wave 1 (see Table 4-4).[10] Overall, these results reflect the use of both demographic variables and tobacco-use measures from Wave 1 for calibrating the Wave 4 interview weights for the Wave 1 Cohort.

Assuming that the Wave 1 demographic, socio-economic, and tobacco-use characteristics in Tables 4-2, 4-3, and 4-4 are correlated with key tobacco and health related outcome measures in Wave 4 of the PATH Study, these results indicate little if any nonresponse bias in the adult interview estimates for the Wave 1 Cohort due to attrition from Wave 1 to Wave 4.

[10] For this particular subgroup, the sample sizes are relatively small and the difference in estimates is 1.5 percentage points.

4.2 Youth Interview

4.2.1 Response Rates Conditioning on Wave 1 Response

Among those in the Wave 1 Cohort, unweighted and weighted response rates were calculated for the Wave 4 youth interview using the following formulas:

$$RR_Y = C_Y/(C_Y + N_Y + e_Y \times U_Y)$$

$$e_Y = (C_Y + N_Y)/(C_Y + N_Y + I_Y)$$

where

RR_Y = Wave 4 youth interview response rate;

C_Y = number of Wave 4 youth interview complete cases;

N_Y = number of Wave 4 youth interview nonrespondents known to be eligible;

U_Y = number of Wave 4 youth interview nonrespondents with unknown eligibility status;

e_Y = estimated proportion of eligible cases among the Wave 4 youth interview nonrespondents with unknown eligibility status; and

I_Y = number of Wave 4 youth interview ineligible cases that were not permanently ineligible prior to Wave 4.

Unweighted and weighted counts based on the Wave 1 IPS weights were obtained for response status categories C_Y, N_Y, U_Y, and I_Y for unweighted response rates and weighted response rates, respectively. A small number of participants became permanently ineligible prior to Wave 4 due to permanent incarceration, death, or having permanently moved overseas; this set of cases, denoted as IP_Y in Table 4-5, were not included in the equation for computing e_Y.

Table 4-5 provides response rates for the Wave 4 youth interview. In addition to the overall row, response rates are shown by response status at earlier waves as well as by Wave 1 sex, age, race/ethnicity, and tobacco-use status. Persons with missing values for a particular characteristic were excluded from the response rate calculation for that characteristic.

Table 4-5. Wave 4 youth interview response rates conditioning on Wave 1 response

Characteristic[a]	C_Y: Completed (n)	IP_Y: Permanently ineligible prior to Wave 4 (n)	I_Y: Other ineligible (n)	N_Y: Nonresponse known to be eligible (n)	U_Y: Nonresponse with unknown eligibility status (n)	RR_Y: Unweighted response rate (%)	RR_Y: Weighted response rate (%)
Overall	11,059	8	36	1,365	1,494	79.5	79.5
Earlier waves' response status							
Responded in all earlier waves	10,433	0	17	633	193	92.7	92.7
Did not respond in at least one earlier wave	626	8	19	732	1,301	23.7	24.1
Wave 1 sex							
Male	5,719	5	20	721	743	79.6	79.8
Female	5,319	2	16	642	743	79.4	79.4
Wave 1 age group							
Under 12	5,452	4	20	760	855	77.2	77.3
12-13	3,808	3	13	410	448	81.6	81.6
14-17	1,799	1	3	195	190	82.4	82.2
Wave 1 race/ethnicity							
Hispanic	3,232	5	18	346	422	80.8	80.9
Non-Hispanic White alone	5,122	1	5	727	731	77.8	77.9
Non-Hispanic Black alone	1,557	2	4	146	177	82.8	82.8
Non-Hispanic other race or multiple races	997	0	9	124	139	79.2	79.0
Wave 1 ever tobacco use[b] (Wave 1 youth only)							
Ever user	528	1	5	69	80	78.1	78.0
Never user	4,807	3	10	494	507	82.8	82.7

[a] For each Wave 1 characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] An 'ever user' is someone who has ever used one or more of the tobacco products covered by the Wave 1 youth interview. A 'never user' is someone who has never used any of those tobacco products. Ever use of a tobacco product is defined as having ever used the product, even one or two times. The products covered by the Wave 1 youth interview are cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, e-cigarettes, dissolvable tobacco, bidis, and kreteks.

The unweighted and weighted response rates for the Wave 4 youth interview were both 79.5 percent, and the unweighted and weighted response rates for each subgroup were also very similar to each other. The response rates varied little by sex, but to a greater extent by Wave 1 age and race/ethnicity. For Wave 1 age, the weighted response rate for the "under 12" age group (77.3 percent) was noticeably lower than for the "12-13" and "14-17" age groups (81.6 percent and 82.2 percent, respectively). This is likely because the persons in the "under 12" age group were shadow youth at Wave 1, so for some of them the Wave 4 youth interview was the first PATH Study interview they were asked to complete. (In Waves 2 and 3, the weighted response rate was also lower among youth who were shadow youth at Wave 1.) Response rates also differed somewhat across race/ethnicity groups, with the highest among non-Hispanic Black alone youth (82.8 percent) and lowest among non-Hispanic White alone youth (77.9 percent). For tobacco use, the weighted response rate among Wave 1 youth was 4.7 percentage points lower for the Wave 1 "ever user" group than for the Wave 1 "never user" group.[11] Similar to the adult interview, the Wave 4 youth interview weighted response rates also differed substantially between those who responded in Waves 2 and 3 (92.7 percent) and those who skipped at least one of these waves (24.1 percent)).

As mentioned in Section 4.1.1, variation in response rates by subgroups is to be expected in large-scale data collection efforts. None of the differences among the demographic and tobacco-use subgroups in Table 4-5 causes serious concern about potential nonresponse bias in the Wave 4 youth interview estimates.

4.2.2 Comparison Between Wave 4 Respondents and Those Eligible for Wave 4 Interview (Among the Wave 1 Cohort)

This section compares weighted estimates between the Wave 4 youth interview respondents in the Wave 1 Cohort and the Wave 1 respondents who were eligible for the Wave 4 youth interview. The weighted estimates cover demographic characteristics (shown in Table 4-6) and tobacco-use status (shown in Table 4-7). The structure of these tables is essentially the same as the tables in Section 4.1.2 for the adult interview.

[11] Cunradi et al. (2005) and Young et al. (2006) have found that smokers were less likely to be retained in subsequent waves of surveys than nonsmokers.

Table 4-6. Comparison of Wave 1 demographic characteristics between Wave 4 youth interview respondents and Wave 1 respondents who were eligible for Wave 4 youth interview

Wave 1 characteristic	Wave 1 respondents who were eligible for Wave 4 youth interview		Wave 1 respondents who completed Wave 4 youth interview				
				Before Wave 4 weighting adjustment		After Wave 4 weighting adjustment	
	Unweighted count	A: Weighted percentage, using Wave 1 final weights [95% confidence interval]	Unweighted count	B: Weighted percentage, using Wave 1 final weights [95% confidence interval]	Difference in weighted percentages [B – A] [95% confidence interval]	C: Weighted percentage, using Wave 4 final weights [95% confidence interval]	Difference in weighted percentages [C – A] [95% confidence interval]
Sex							
Male	7,183	51.3% [50.4%, 52.1%]	5,719	51.4% [50.5%, 52.4%]	0.1% [-0.4%, 0.7%]	51.3% [50.4%, 52.2%]	0.0% [-0.1%, 0.1%]
Female	6,704	48.7% [47.9%, 49.6%]	5,319	48.6% [47.6%, 49.5%]	-0.1% [-0.7%, 0.4%]	48.7% [47.8%, 49.6%]	0.0% [-0.1%, 0.1%]
Age group							
Under 12	7,067	49.5% [48.7%, 50.3%]	5,452	48.1% [47.2%, 49.1%]	-1.4% [-1.8%, -0.9%]	49.1% [48.2%, 50.0%]	-0.4% [-0.5%, -0.3%]
12-13	4,666	34.4% [33.6%, 35.2%]	3,808	35.2% [34.3%, 36.1%]	0.8% [0.4%, 1.2%]	34.2% [33.4%, 35.1%]	-0.2% [-0.2%, -0.1%]
14-17	2,184	16.1% [15.5%, 16.7%]	1,799	16.6% [16.0%, 17.4%]	0.5% [0.2%, 0.9%]	16.7% [16.0%, 17.4%]	0.6% [0.4%, 0.7%]
Race/ethnicity							
Hispanic	4,000	23.4% [22.7%, 24.1%]	3,232	23.7% [22.9%, 24.5%]	0.3% [-0.2%, 0.8%]	23.3% [22.5%, 24.1%]	-0.1% [-0.2%, 0.0%]
Non-Hispanic White alone	6,580	53.2% [52.3%, 54.0%]	5,122	52.3% [51.4%, 53.3%]	-0.8% [-1.4%, -0.3%]	53.2% [52.3%, 54.2%]	0.1% [-0.0%, 0.2%]
Non-Hispanic Black alone	1,880	13.9% [13.3%, 14.5%]	1,557	14.4% [13.7%, 15.0%]	0.5% [0.1%, 0.8%]	13.9% [13.2%, 14.5%]	0.0% [-0.1%, 0.1%]
Non-Hispanic other race or multiple races	1,260	9.6% [9.1%, 10.1%]	997	9.6% [9.1%, 10.2%]	0.0% [-0.3%, 0.3%]	9.6% [9.1%, 10.2%]	0.0% [-0.1%, 0.1%]

Table 4-7. Comparison of Wave 1 "ever tobacco-use" estimates between Wave 1 youth who completed Wave 4 youth interview and Wave 1 youth who were eligible for Wave 4 youth interview*

Wave 1 characteristic[a]	Wave 1 youth who were eligible for Wave 4 youth interview		Wave 4 youth interview respondents who completed Wave 1 youth interview				
				Before Wave 4 weighting adjustment		After Wave 4 weighting adjustment	
	Sample size	A: Weighted estimate, using Wave 1 final weights [95% confidence interval]	Sample size	B: Weighted estimate, using Wave 1 final weights [95% confidence interval]	Difference in weighted estimates [B – A] [95% confidence interval]	C: Weighted estimate, using Wave 4 final weights [95% confidence interval]	Difference in weighted estimates [C – A] [95% confidence interval]
Overall	6,485	10.1% [9.3%, 11.0%]	5,335	9.5% [8.6%, 10.6%]	-0.6% [-1.0%, -0.1%]	9.6% [8.6%, 10.6%]	-0.5% [-1.0%, -0.1%]
Sex							
Male	3,285	11.2% [10.1%, 12.5%]	2,710	10.7% [9.4%, 12.1%]	-0.6% [-1.2%, 0.0%]	10.8% [9.5%, 12.2%]	-0.5% [-1.1%, 0.1%]
Female	3,176	8.9% [7.8%, 10.1%]	2,608	8.4% [7.2%, 9.8%]	-0.5% [-1.1%, 0.1%]	8.3% [7.1%, 9.7%]	-0.5% [-1.1%, 0.0%]
Age group[b]							
12-13	4,385	7.8% [6.9%, 8.8%]	3,597	7.4% [6.4%, 8.6%]	-0.4% [-0.8%, 0.1%]	7.5% [6.5%, 8.6%]	-0.3% [-0.8%, 0.1%]
14-17	2,100	14.9% [13.4%, 16.5%]	1,738	13.8% [12.2%, 15.6%]	-1.0% [-1.9%, -0.1%]	13.8% [12.2%, 15.6%]	-1.1% [-2.0%, -0.1%]
Race/ethnicity							
Hispanic	1,894	9.5% [8.3%, 10.9%]	1,592	9.3% [7.9%, 10.8%]	-0.2% [-0.8%, 0.4%]	9.3% [7.9%, 10.8%]	-0.2% [-0.8%, 0.4%]
Non-Hispanic White alone	2,974	10.6% [9.4%, 11.9%]	2,399	9.7% [8.3%, 11.3%]	-0.9% [-1.7%, -0.1%]	9.7% [8.3%, 11.3%]	-0.9% [-1.7%, -0.1%]
Non-Hispanic Black alone	862	9.6% [7.5%, 12.3%]	730	9.5% [6.9%, 12.8%]	-0.2% [-1.2%, 0.8%]	9.6% [7.1%, 12.9%]	0.0% [-1.0%, 1.0%]
Non-Hispanic other race or multiple races	590	9.5% [7.4%, 12.2%]	483	9.3% [7.0%, 12.2%]	-0.2% [-1.4%, 1.0%]	9.4% [7.1%, 12.3%]	-0.1% [-1.2%, 1.0%]

* An 'ever user' is someone who has ever used one or more of the tobacco products covered by the Wave 1 youth interview. A 'never user' is someone who has never used any of those tobacco products. Ever use of a tobacco product is defined as having ever used the product, even one or two times. The products covered by the Wave 1 youth interview are cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, e-cigarettes, dissolvable tobacco, bidis, and kreteks.

[a] For each Wave 1 characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] Table 4-7 includes a subset of the youth in Table 4-6 because there are no Wave 1 interview data for the Wave 4 youth interview respondents who were shadow youth at Wave 1.

Table 4-6 compares Wave 1 demographic characteristics between the Wave 4 youth interview respondents and those eligible for the Wave 4 youth interview. In Table 4-6, as in Table 4-2, the percentages sum to 100 percent[12] over the categories associated with each characteristic. For sex, the "before Wave 4 weighting adjustment" distributions among the Wave 4 youth interview respondents were almost the same as the distributions among those eligible for the Wave 4 youth interview. Non-Hispanic White alone youth were slightly underrepresented among the Wave 4 respondents and non-Hispanic Black alone youth were slightly overrepresented. The most noticeable underrepresentation among the Wave 4 youth respondents was the Wave 1 "under 12" age group (with corresponding slight overrepresentation among the two older age groups).

For the "after Wave 4 weighting adjustment" differences, the estimated confidence intervals (in the last column of Table 4-6) either included zero or were in the proximity of zero for sex and race/ethnicity. For Wave 1 age, there is a slight underrepresentation of the "under 12" and "12-13" age groups, with a corresponding slight overrepresentation of the "14-17" age group.

Table 4-7 shows the comparison of Wave 1 "ever tobacco-use" rates between Wave 1 youth interview respondents who completed the Wave 4 youth interview and all the Wave 1 youth interview respondents eligible for the Wave 4 youth interview. (Table 4-7 includes a subset of the youth in Table 4-6 because there are no Wave 1 interview data for the Wave 4 youth interview respondents who were shadow youth at Wave 1.) For both the "before Wave 4 weighting adjustment" and the "after Wave 4 weighting adjustment" estimates, differences between respondents and those eligible for the interview were no more than 1.1%. The largest difference was for the Wave 1 "ever tobacco use" measure among youth ages 14-17 at that time. However, most of the confidence intervals either included or were in the proximity of zero.

Note that only the Wave 1 14-year-olds among the Wave 1 "14-17" age group would be expected to be youth at Wave 4, so these results may in part reflect the fact that not all study participants were exactly 3 years older at the time of their Wave 4 participation, compared to their Wave 1 age.

Assuming that the Wave 1 demographic and tobacco-use characteristics in Tables 4-6 and 4-7 are correlated with key tobacco and health related outcome measures in Wave 4 of the PATH Study,

[12] The sum may be not exactly 100 percent due to rounding.

these results indicate little if any nonresponse bias in the youth interview estimates for the Wave 1 Cohort due to attrition from Wave 1 to Wave 4.

4.3 Adult Biospecimen Collections

4.3.1 Unweighted Response Rates

This section discusses unweighted response rates for the urine specimen collection and blood specimen collection among Wave 4 adult interview respondents who belong to the Wave 1 Cohort and were asked to provide a biospecimen.[13] Each response rate was calculated as a ratio, with the numerator being the number of Wave 1 Cohort adults who provided a biospecimen and the denominator being the number of Wave 1 Cohort adults from whom a biospecimen was requested.

Tables 4-8 and 4-9 show the unweighted response rates for urine collection and blood collection, respectively. In addition to the overall rows, the response rates are tabulated by Wave 4 socio-demographic characteristics and tobacco-use status for both urine and blood collection. For urine collection, response rates are also presented for subgroups defined by whether the respondent is a member of the Wave 1 Biomarker Core, and whether Wave 4 was the first time the respondent had completed a PATH Study adult interview. Adults with missing values for a particular characteristic were excluded from the response rate calculation for that characteristic.

As shown in Table 4-8, the unweighted response rate for urine collection was 96.5 percent overall, with little variation by sex, race/ethnicity, education, and health insurance status. The urine collection response rates were higher for adults in the two middle age groups (98.3 percent for ages 25-44 and 97.9 percent for ages 45-64) than those in the younger or older age groups (94.2 percent for ages 18-24 and 95.8 percent for ages 65 and above). Current established tobacco users at Wave 4 were more likely to provide urine specimens (98.1 percent) than adults who were not current established users (95.2 percent). First-time adult interview respondents were less likely to provide urine specimens, and Wave 1 Biomarker Core members were more likely to provide urine

[13] Weighted response rates are not provided because the subset of Wave 1 adults asked to provide a urine specimen at Wave 4 does not represent a readily interpretable portion of the population. Similarly, while most first-time Adult Interview respondents were age 18 at Wave 4, some were older due to the actual time elapsed between the Wave 1 and Wave 4 interviews.

specimens. Despite the differences in response propensities across subgroups, the response rates for all subgroups examined were above 87 percent.

Table 4-8. Unweighted urine collection response rates among Wave 4 adult interview respondents selected to provide a urine specimen (Wave 1 Cohort)

Wave 4 characteristic[a] or Biomarker Core membership	Wave 4 adult interview respondents selected to provide urine (n)	Urine providers among Wave 4 adult interview respondents (n)	Unweighted response rate for urine collection (%)
Overall	16,711	16,131	96.5
Sex			
Male	8,349	8,032	96.2
Female	8,349	8,086	96.8
Age group			
18-24	6,237	5,878	94.2
25-44	5,695	5,597	98.3
45-64	3,692	3,615	97.9
65 and above	1,087	1,041	95.8
Race/ethnicity			
Hispanic	3,423	3,298	96.3
Non-Hispanic White alone	9,202	8,852	96.2
Non-Hispanic Black alone	2,599	2,539	97.7
Non-Hispanic other race or multiple races	1,262	1,222	96.8
Education			
Less than high school or GED	3,618	3,497	96.7
High school	4,424	4,222	95.4
Some college, no degree	6,005	5,833	97.1
Bachelor degree and above	2,590	2,509	96.9
Health insurance			
Yes	13,432	12,958	96.5
No	3,108	3,018	97.1
Tobacco-use status[b]			
Current established user	7,628	7,486	98.1
Not current established user	9,035	8,598	95.2
First time completing adult interview			
Yes	1,900	1,660	87.4
No	14,811	14,471	97.7
Wave 1 Biomarker Core member			
Yes	8,807	8,641	98.1
No	7,904	7,490	94.8

[a] All the characteristics apply to the Wave 4 adult interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth or adult interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] A tobacco user is defined as someone who uses one or more of the following tobacco products covered by the Wave 4 adult interview: cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, and electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah). A 'current established user' of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly.

Table 4-9. Unweighted blood collection response rates among Wave 4 first-time adult interview respondents (Wave 1 Cohort)

Wave 4 characteristic[a]	First-time adult interview respondents at Wave 4 (n)	Blood providers among first-time adult interview respondents at Wave 4 (n)	Unweighted response rate for blood collection (%)
Overall	1,900	890	46.8
Sex			
Male	988	439	44.4
Female	910	449	49.3
Race/ethnicity			
Hispanic	556	299	53.8
Non-Hispanic White alone	900	391	43.4
Non-Hispanic Black alone	246	105	42.7
Non-Hispanic other race or multiple races	168	81	48.2
Education			
Less than high school or GED	622	301	48.4
High school	925	426	46.1
Some college, no degree	324	149	46.0
Bachelor degree and above	10	5	50.0
Health insurance			
Yes	1,483	712	48.0
No	367	160	43.6
Tobacco-use status[b]			
Current established user	251	113	45.0
Not current established user	1,644	775	47.1

[a] All the characteristics apply to the Wave 4 adult interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth or adult interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] A tobacco user is defined as someone who uses one or more of the following tobacco products covered by the Wave 4 adult interview: cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, and electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah). A 'current established user' of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly.

As shown in Table 4-9, the unweighted response rate for blood collection among first-time adult interview respondents was 46.8 percent. Females (49.3 percent) were more likely to provide blood specimens than males (44.4 percent). The biggest subgroup difference was for race/ethnicity, with a 53.8 percent response rate for Hispanic adults and response rates ranging from 42.7 percent to 48.2 percent for non-Hispanic adults. Also, blood collection response rates were lower for adults without health insurance and for current established tobacco users. The modest overall response rate for blood collection combined with moderate variation in response propensities by subgroups indicates some potential for nonresponse bias with respect to these data.

4.3.2 Comparison Between First-Time Adult Interview Respondents Who Provided a Biospecimen at Wave 4 and All First-Time Adult Interview Respondents (Among the Wave 1 Cohort)

All first-time adult interview respondents at Wave 4 were asked to provide a urine specimen and a blood specimen. Table 4-10 compares the characteristics of Wave 4 first-time adult interview respondents who provided a urine specimen with all Wave 4 first-time adult interview respondents, among the Wave 1 Cohort. Table 4-11 presents a similar analysis for the collection of blood specimens. The Wave 4 single-wave weights (which account for Wave 4 interview nonresponse) were used for generating all the estimates.

Despite the nominal differences for some estimates, most of the 95 percent confidence intervals for differences in Tables 4-10 and 4-11 included zero. The exceptions involved race/ethnicity subgroups for both the urine and blood collections, and a difference in the sex distribution for the blood collection only. Among the urine providers, there was a slight underrepresentation of non-Hispanic White alone adults and a slight overrepresentation of non-Hispanic Black alone adults. Among the blood providers, males and non-Hispanic White alone adults were underrepresented and Hispanic adults were overrepresented to some extent.

Note that this analysis applies to first-time adult interview respondents only, not to all Wave 4 adult interview respondents in the Wave 1 Cohort. Among this cohort, most first-time adult interview respondents at Wave 4 were 18 years old, but a small proportion may have been older. The plan for laboratory analysis of Wave 4 adult biospecimens is unknown at this time, so the results in Tables 4-10 and 4-11 may not correspond to those who will have Wave 4 biomarker data. Furthermore, a set of biospecimen weights may be created for analyzing the adult biomarker data, and those weights would attempt to correct for potential nonresponse bias associated with differential biospecimen collection response rates.

Table 4-10. Comparison of characteristics between Wave 4 first-time adult interview respondents who provided a urine specimen and all Wave 4 first-time adult interview respondents (Wave 1 Cohort)

Wave 4 characteristic[a]	First-time adult interview respondents at Wave 4		Urine providers among first-time adult interview respondents at Wave 4		Difference in weighted percentages [urine providers – first-time adult interview respondents at Wave 4] [95% confidence interval]
	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	
Sex					
Male	988	51.4% [49.2%, 53.6%]	868	51.3% [48.9%, 53.7%]	-0.1% [-1.1%, 0.9%]
Female	910	48.6% [46.4%, 50.8%]	790	48.7% [46.3%, 51.1%]	0.1% [-0.9%, 1.1%]
Race/ethnicity					
Hispanic	556	22.4% [20.5%, 24.3%]	494	23.0% [21.0%, 25.1%]	0.6% [-0.1%, 1.4%]
Non-Hispanic White alone	900	56.8% [54.5%, 59.0%]	764	55.6% [53.1%, 58.0%]	-1.2% [-2.3%, -0.2%]
Non-Hispanic Black alone	246	11.7% [10.4%, 13.3%]	227	12.5% [10.9%, 14.1%]	0.7% [0.1%, 1.4%]
Non-Hispanic other race or multiple races	168	9.1% [7.9%, 10.5%]	147	9.0% [7.7%, 10.4%]	-0.1% [-0.9%, 0.6%]
Education					
Less than high school or GED	622	33.1% [31.0%, 35.3%]	554	33.3% [31.1%, 35.7%]	0.2% [-0.8%, 1.3%]
High school	925	48.3% [45.9%, 50.7%]	799	47.9% [45.2%, 50.5%]	-0.4% [-1.6%, 0.7%]
Some college, no degree	324	18.1% [16.2%, 20.2%]	282	18.3% [16.3%, 20.5%]	0.2% [-0.6%, 1.0%]
Bachelor degree and above	10	0.5% [0.2%, 0.9%]	9	0.5% [0.3%, 1.0%]	0.0% [-0.0%, 0.1%]

Table 4-10. Comparison of characteristics between Wave 4 first-time adult interview respondents who provided a urine specimen and all Wave 4 first-time adult interview respondents (Wave 1 Cohort) (continued)

Wave 4 characteristic[a]	First-time adult interview respondents at Wave 4		Urine providers among first-time adult interview respondents at Wave 4		Difference in weighted percentages [urine providers – first-time adult interview respondents at Wave 4] [95% confidence interval]
	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	
Health insurance					
Yes	1,483	81.7% [79.7%, 83.6%]	1,296	81.4% [79.4%, 83.3%]	-0.3% [-1.0%, 0.3%]
No	367	18.3% [16.4%, 20.3%]	325	18.6% [16.7%, 20.6%]	0.3% [-0.3%, 1.0%]
Tobacco-use status[b]					
Current established user	251	14.2% [12.5%, 16.1%]	228	14.6% [12.7%, 16.7%]	0.4% [-0.4%, 1.2%]
Not current established user	1,644	85.8% [83.9%, 87.5%]	1,428	85.4% [83.3%, 87.3%]	-0.4% [-1.2%, 0.4%]

[a] All the characteristics apply to the Wave 4 adult interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth or adult interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] A tobacco user is defined as someone who uses one or more of the following tobacco products covered by the Wave 4 adult interview: cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, and electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah). A 'current established user' of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly.

Table 4-11. Comparison of characteristics between Wave 4 first-time adult interview respondents who provided a blood specimen and all Wave 4 first-time adult interview respondents (Wave 1 Cohort)

Wave 4 characteristic[a]	First-time adult interview respondents at Wave 4		Blood providers among first-time adult interview respondents at Wave 4		Difference in weighted percentages [blood providers – first-time adult interview respondents at Wave 4] [95% confidence interval]
	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	
Sex					
Male	988	51.4% [49.2%, 53.6%]	439	48.0% [44.7%, 51.2%]	-3.4% [-5.7%, -1.2%]
Female	910	48.6% [46.4%, 50.8%]	449	52.0% [48.8%, 55.3%]	3.4% [1.2%, 5.7%]
Race/ethnicity					
Hispanic	556	22.4% [20.5%, 24.3%]	299	26.6% [23.8%, 29.6%]	4.2% [1.9%, 6.5%]
Non-Hispanic White alone	900	56.8% [54.5%, 59.0%]	391	53.0% [49.5%, 56.4%]	-3.8% [-6.7%, -1.0%]
Non-Hispanic Black alone	246	11.7% [10.4%, 13.3%]	105	10.8% [8.9%, 13.2%]	-0.9% [-2.8%, 1.0%]
Non-Hispanic other race or multiple races	168	9.1% [7.9%, 10.5%]	81	9.6% [7.8%, 11.8%]	0.5% [-1.3%, 2.4%]
Education					
Less than high school or GED	622	33.1% [31.0%, 35.3%]	301	33.6% [30.5%, 36.8%]	0.5% [-2.2%, 3.1%]
High school	925	48.3% [45.9%, 50.7%]	426	47.5% [44.2%, 50.9%]	-0.8% [-3.3%, 1.7%]
Some college, no degree	324	18.1% [16.2%, 20.2%]	149	18.4% [15.6%, 21.5%]	0.3% [-1.8%, 2.3%]
Bachelor degree and above	10	0.5% [0.2%, 0.9%]	5	0.5% [0.2%, 1.3%]	0.1% [-0.2%, 0.4%]

Table 4-11. Comparison of characteristics between Wave 4 first-time adult interview respondents who provided a blood specimen and all Wave 4 first-time adult interview respondents (Wave 1 Cohort) (continued)

Wave 4 characteristic[a]	First-time adult interview respondents at Wave 4		Blood providers among first-time adult interview respondents at Wave 4		Difference in weighted percentages [blood providers – first-time adult interview respondents at Wave 4] [95% confidence interval]
	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	
Health insurance					
Yes	1,483	81.7% [79.7%, 83.6%]	712	83.2% [80.0%, 85.9%]	1.4% [-0.6%, 3.5%]
No	367	18.3% [16.4%, 20.3%]	160	16.8% [14.1%, 20.0%]	-1.4% [-3.5%, 0.6%]
Tobacco-use status[b]					
Current established user	251	14.2% [12.5%, 16.1%]	113	13.5% [11.0%, 16.5%]	-0.7% [-2.7%, 1.4%]
Not current established user	1,644	85.8% [83.9%, 87.5%]	775	86.5% [83.5%, 89.0%]	0.7% [-1.4%, 2.7%]

[a] All the characteristics apply to the Wave 4 adult interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth or adult interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] A tobacco user is defined as someone who uses one or more of the following tobacco products covered by the Wave 4 adult interview: cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, and electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah). A 'current established user' of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly.

4.4 Youth Urine Specimen Collection

4.4.1 Unweighted Response Rates

This section discusses unweighted response rates for the urine specimen collection among Wave 4 youth interview respondents who belong to the Wave 1 Cohort, all of whom were asked to provide a urine specimen. Each response rate was calculated as a ratio, with the numerator being the number of Wave 1 Cohort youth who provided a urine specimen and the denominator being the number of Wave 4 youth interview respondents in the Wave 1 Cohort.

Table 4-12 shows the unweighted response rates for youth urine collection. In addition to the overall row, the response rates are tabulated by Wave 4 demographic characteristics and tobacco-use status. Youth with missing values for a particular characteristic were excluded from the response rate calculation for that characteristic.

Table 4-12. Unweighted urine collection response rates among Wave 4 youth interview respondents (Wave 1 Cohort)

Wave 4 characteristic[a]	Wave 4 youth interview respondents (n)	Urine providers among Wave 4 youth interview respondents (n)	Unweighted response rate for urine collection (%)
Overall	11,059	9,892	89.4
Sex			
Male	5,720	5,266	92.1
Female	5,301	4,594	86.7
Age group			
12-13	3,482	3,060	87.9
14-17	7,577	6,832	90.2
Race/ethnicity			
Hispanic	3,214	2,938	91.4
Non-Hispanic White alone	4,902	4,282	87.4
Non-Hispanic Black alone	1,452	1,356	93.4
Non-Hispanic other race or multiple races	1,015	902	88.9
Tobacco-use status[b]			
Ever user	1,091	1,032	94.6
Never user	9,650	8,585	89.0

[a] All the characteristics apply to the Wave 4 youth interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] An 'ever user' is someone who has ever used one or more of the tobacco products covered by the Wave 4 youth interview. A 'never user' is someone who has never used any of those tobacco products. Ever use of a tobacco product is defined as having ever used the product, even one or two times. The products covered by the Wave 4 youth interview are cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah), dissolvable tobacco, bidis, and kreteks.

The unweighted response rate for youth urine collection was 89.4 percent overall, with little variation by age. Males (92.1 percent) and youth who had ever used tobacco (94.6 percent) were more likely to provide a urine specimen than females (86.7 percent) and never users of tobacco (89.0 percent). The response rates varied by 6 percentage points across race/ethnicity groups, with non-Hispanic Black alone youth being the most likely to provide a specimen (93.4 percent).

4.4.2 Comparison Between Youth Interview Respondents Who Provided a Urine Specimen at Wave 4 and All Youth Interview Respondents (Among the Wave 1 Cohort)

All youth interview respondents in Wave 4 were asked to provide a urine specimen. Table 4-13 compares the characteristics of Wave 4 youth interview respondents who provided a urine specimen with all Wave 4 youth interview respondents, among the Wave 1 Cohort. The Wave 4 single-wave weights were used for generating all the estimates.

Consistent with the findings about subgroup response rates, Table 4-13 shows that compared to those youth respondents asked to a provide urine specimen, the urine providers slightly overrepresent males, older youth (ages 14-17), Hispanic and non-Hispanic Black alone youth, and youth who had ever used tobacco, and slightly underrepresent non-Hispanic White alone youth. Although the subgroup differences were statistically significant for these characteristics, all the differences were below two percentage points.

The plan for laboratory analysis of Wave 4 youth urine specimens is unknown at this time, so the results in Table 4-13 may not correspond to those who will have Wave 4 biomarker data. A set of biospecimen weights may be created for analyzing the youth biomarker data, and those weights would attempt to correct for potential nonresponse bias associated with differential urine collection response rates.

Table 4-13. Comparison of characteristics between Wave 4 youth interview respondents who provided a urine specimen and all Wave 4 youth interview respondents (Wave 1 Cohort)

Wave 4 characteristic[a]	Wave 4 youth interview respondents		Urine providers among Wave 4 youth interview respondents		Difference in weighted percentages [urine providers – Wave 4 youth interview respondents] [95% confidence interval]
	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	
Sex					
Male	5,720	51.5% [50.5%, 52.4%]	5,266	53.0% [52.0%, 54.0%]	1.6% [1.2%, 2.0%]
Female	5,301	48.5% [47.6%, 49.5%]	4,594	47.0% [46.0%, 48.0%]	-1.6% [-2.0%, -1.2%]
Age group					
12-13	3,482	31.9% [31.0%, 32.7%]	3,060	31.1% [30.2%, 32.1%]	-0.7% [-1.0%, -0.4%]
14-17	7,577	68.1% [67.3%, 69.0%]	6,832	68.9% [67.9%, 69.8%]	0.7% [0.4%, 1.0%]
Race/ethnicity					
Hispanic	3,214	24.2% [23.4%, 25.0%]	2,938	24.7% [23.9%, 25.6%]	0.5% [0.2%, 0.8%]
Non-Hispanic White alone	4,902	52.3% [51.4%, 53.3%]	4,282	51.3% [50.3%, 52.3%]	-1.0% [-1.4%, -0.6%]
Non-Hispanic Black alone	1,452	13.3% [12.6%, 13.9%]	1,356	13.9% [13.2%, 14.6%]	0.6% [0.3%, 0.9%]
Non-Hispanic other race or multiple races	1,015	10.2% [9.6%, 10.8%]	902	10.1% [9.5%, 10.7%]	-0.1% [-0.4%, 0.1%]
Tobacco-use status[b]					
Ever user	1,091	10.1% [9.5%, 10.8%]	1,032	10.7% [10.0%, 11.5%]	0.6% [0.4%, 0.8%]
Never user	9,650	89.9% [89.2%, 90.5%]	8,585	89.3% [88.5%, 90.0%]	-0.6% [-0.8%, -0.4%]

[a] All the characteristics apply to the Wave 4 youth interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] An 'ever user' is someone who has ever used one or more of the tobacco products covered by the Wave 4 youth interview. A 'never user' is someone who has never used any of those tobacco products. Ever use of a tobacco product is defined as having ever used the product, even one or two times. The products covered by the Wave 4 youth interview are cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah), dissolvable tobacco, bidis, and kreteks.

5. Results of Wave 4 Cohort Nonresponse Bias Analyses

This chapter presents results of nonresponse bias analyses for the Wave 4 Cohort. Sections 5.1.1, 5.2.1, and 5.3.1, respectively, describe how the household screener, adult interview, and youth interview response rates vary across select subgroups among the AYS replenishment sample. Sections 5.2.2 and 5.3.2 compare cross-sectional estimates of demographic and socio-economic characteristics as well as cigarette-use status between the PATH Study Wave 4 Cohort and external sources. Sections 5.4 and 5.5 focus on the analyses of biospecimen data collections. All results are presented separately for adults and youth.

As discussed in Section 3.2.1, 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. As demonstrated by the results in Tables 4-1 and 4-5 for the Wave 1 Cohort, differences in response status prior to Wave 4 led to large differences in Wave 4 response rates. Furthermore, Wave 4 response rates are conditioning on Wave 1 response for those sampled at Wave 1, whereas Wave 4 response rates for those from the replenishment sample 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. Instead, response rates were calculated separately for the Wave 1 Cohort[14] and the Wave 4 replenishment sample. The Wave 4 adult interview and youth interview response rates among Wave 1 respondents are presented in Sections 4.1.1 and 4.2.1, respectively. This chapter therefore focuses on the interview response rates calculated for adults and youth from the Wave 4 replenishment sample, and the household screener response rates calculated for the AYS replenishment sample (from which all Wave 4 replenishment sample adults and youth were selected) at Wave 4.

[14] 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.

5.1 Household Screener

5.1.1 Household Screener Response Rates for Wave 4 Adult, Youth, and Shadow Youth (AYS) Replenishment Sample

For the AYS replenishment sample, the household screener combined typical screener functions (e.g., enumerating the household, collecting demographic information about each member and some household-level data, and selecting participants for the study) with a special purpose for the PATH Study, collecting basic information on each adult's tobacco use. This allowed the study to classify the adult with sufficient validity for potential selection as a participant based on the PATH Study's sampling strata, even though some degree of misreporting of tobacco-use and demographic characteristics by the household screener respondent was expected.

The household screener response rate was calculated as

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

where

R_{HH} = Wave 4 AYS replenishment sample household screener response rate;

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.

Unweighted counts and weighted counts based on the Wave 4 household IPS weights were obtained for each response status category, for unweighted response rates and weighted response rates, respectively. Table 5-1 provides overall response rates for the household screener and response rates by subgroups of sampled households. Subgroups were defined by the characteristics of the census block groups in which the sampled addresses were located, using information from the 5-year (2012 to 2016) ACS.[15] The "high" and "low" subgroup categories were defined relative to the nationwide

[15] Information from the 5-year (2012 to 2016) rather than the 1-year (2016) ACS was used because 1-year ACS estimates are not provided for smaller geographies such as census tracts or block groups.

percentage of persons having the characteristic: block groups with percentages below the national average for the characteristic were classified as low, and those with percentages above the national average were classified as high.

As shown in Table 5-1, the weighted overall response rate for the AYS replenishment sample household screener was 52.8 percent. The weighted response rates of subgroups differed from one another by 9.9 percentage points for education, 7.5 percentage points for poverty status, and 6.5 percent by census region. The weighted subgroup response rates varied no more than 1.5 percentage points for race or ethnicity.

Table 5-1. Wave 4 household screener response rates for adult, youth, and shadow youth (AYS) replenishment sample

Neighborhood characteristic[b] or census region	C_{HH}: Completed (n)	N_{HH}: Nonresponse known to be eligible (n)	$e_{HH} \times U_{HH}$: Nonresponse with unknown eligibility estimated to be eligible (n)	RR_{HH}: Unweighted response rate[a] (%)	RR_{HH}: Weighted response rate[a] (%)
Overall	27,364	22,010	2,230	53.0	52.8
Education (% of persons ages 25+ with bachelor's degree)					
High (> 30.3%)	10,841	10,480	1,498	47.5	47.3
Low (<= 30.3%)	16,523	11,530	733	57.4	57.2
Race (% Black alone)					
High (> 12.6%)	6,955	5,277	618	54.1	53.8
Low (<= 12.6%)	20,409	16,733	1,612	52.7	52.5
Race (% White alone)					
High (> 73.3%)	18,138	14,551	1,239	53.5	53.3
Low (<= 73.3%)	9,226	7,459	991	52.2	51.8
Ethnicity (% Hispanic)					
High (> 17.3%)	7,484	6,081	679	52.5	52.3
Low (<= 17.3%)	19,880	15,929	1,552	53.2	53.0
Poverty status (% below poverty line)					
High (> 15.1%)	10,843	7,147	773	57.8	57.6
Low (<= 15.1%)	16,521	14,863	1,457	50.3	50.1
Census region					
Northeast	4660	4228	379	50.3	50.0
Midwest	6556	4665	372	56.6	56.5
South	10,248	8,154	905	53.1	52.9
West	5,900	4,963	575	51.6	51.4

[a] The unweighted response rate = $C_{HH}/(C_{HH}+N_{HH}+e_{HH}*U_{HH})$. The weighted response rate was calculated using a similar approach except that the Wave 4 household IPS weights were used to obtain weighted counts.

[b] The information used to define the subgroups was from the 5-year (2012-2016) ACS.

5.2 Adult Interview

5.2.1 Response Rates for Wave 4 Replenishment Sample

For the Wave 4 replenishment sample, the adult interview response rate (conditioning on completion of the household screener) was calculated as the product of (1) the Phase 2 or individual 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. The response rate was calculated as:

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

where

RR_A = Wave 4 replenishment sample adult interview response rate;

C_{P2} = number of Phase 2 screener complete cases;

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

C_A = number of Wave 4 adult interview complete cases; and

N_A = number of Wave 4 adult interview nonrespondents.

Unweighted response rates were calculated based on actual case counts, while weighted response rates were based on the sums of the Wave 4 IPS weights (calculated as the product of the household IPS weight and the inverse of the within-household probability of selection).

Table 5-2 provides the adult interview response rates. In addition to the overall row, the table includes rows on age, sex, race/ethnicity, and tobacco-use status based on information collected through the household screener, as well as census region. Among adults for whom the household screener had missing values for information on tobacco-use status, sampling at Phase 1 was based on the selection probabilities used for tobacco users, shown in the "sampled as user" row of Table 5-2. Adults with missing values for other characteristics were excluded from the response rate calculation for those characteristics.

The weighted overall response rate for the adult interview was 68.0 percent. The biggest subgroup difference occurred for race/ethnicity, with weighted response rates of 71.7 percent for non-Hispanic Black alone adults and 61.0 percent for non-Hispanic adults of other or multiple races. The

differences among subgroups on weighted response rates were 3.5 percentage points for tobacco-use status, 2.8 percentage points for age, 5.6 percentage points for sex, and 7.4 percentage points for census region.

Table 5-2. Wave 4 adult interview response rates conditioning on household screener response (Wave 4 replenishment sample)

Characteristic reported by household screener respondent[a] or census region	Phase 2 screener		Adult interview		RR_A: Unweighted response rate[b] (%)	RR_A: Weighted response rate[b] (%)
	C_{P2}: P2 screener completed (n)	N_{P2}: P2 screener non-response (n)	C_A: Adult interview completed (n)	N_A: Adult interview non-response (n)		
Overall	8,927	4,120	6,065	12	68.3	68.0
Tobacco-use status						
Sampled as user	4,404	1,750	3,786	8	71.4	70.2
Sampled as nonuser	4,523	2,370	2,279	4	65.5	66.7
Age group						
18-24	3,193	1,573	2,619	2	66.9	66.6
25-44	2,416	1,077	1,693	1	69.1	68.4
45-64	2,113	982	1,235	5	68.0	67.3
65 and above	1,149	488	482	4	69.6	69.4
Sex						
Male	4,316	2,246	3,135	8	65.6	65.1
Female	4,611	1,873	2,930	4	71.0	70.7
Race/ethnicity						
Hispanic	1,809	770	1,197	0	70.1	70.3
Non-Hispanic White alone	5,024	2,431	3,344	7	67.3	67.7
Non-Hispanic Black alone	1,335	494	992	5	72.6	71.7
Non-Hispanic other race or multiple races	737	412	518	0	64.1	61.0
Census region						
Northeast	1,333	792	881	4	62.4	62.5
Midwest	2,103	897	1,427	2	70.0	69.9
South	3,468	1,559	2,398	5	68.8	68.3
West	2,023	872	1,359	1	69.8	69.7

[a] For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values or unusual circumstances. For example, the sum of the counts across the age categories is not equal to the count in the overall row because a small proportion of cases were reported by the household screener respondent as under age 18.

[b] The unweighted response rate = $(C_{P2}/(C_{P2}+N_{P2}))*(C_A/(C_A+N_A))$. The weighted response rate was calculated using a similar approach except that the Wave 4 IPS weights were used to obtain weighted counts.

5.2.2 Comparison of Wave 4 Cohort Socio-Demographic Estimates to the American Community Survey

This nonresponse bias analysis investigates possible differences between weighted socio-demographic estimates based on the Wave 4 Cohort of the PATH Study and corresponding estimates from the ACS. Although the response rates in Sections 5.1.1 and 5.2.1 pertain to the AYS replenishment sample, the analyses in this section were conducted on the Wave 4 Cohort. The 1-year (2016) ACS estimates, calculated from the 2016 ACS PUMS,[16] were used for comparison purposes. Estimates calculated from the ACS PUMS excluded institutional group quarters and persons in noninstitutional group quarters who were not college students. These exclusions correspond to the PATH Study's exclusions from the Wave 4 Cohort target population. By so doing, the PATH Study is able to assess the extent to which differential nonresponse among population subgroups may affect Wave 4 Cohort estimates. Estimates for the PATH Study were calculated using the Wave 4 Cohort cross-sectional weights. In this section, results are presented for adults. Section 5.3.2 presents similar results for youth.

Table 5-3 presents weighted estimates of adult socio-demographic characteristics based on the Wave 4 Cohort adults. The weighted percentages of sex, age, race/ethnicity, and education are a close match to the 1-year 2016 ACS estimates of those quantities. The Wave 4 Cohort cross-sectional weights were calibrated to match the 2016 ACS totals for these four characteristics, so this close match is the expected and desired result. Table 5-3 also shows that the estimated percentage of adults with health insurance based on the Wave 4 Cohort (88.1 percent) is 3.8 percentage points lower than the corresponding estimate from the 2016 ACS (91.9 percent). In Table 5-3, the adults with missing values for a particular characteristic were excluded from the Wave 4 Cohort estimates for that characteristic.[17]

[16] The ACS PUMS files provide a probability sample of individual records from the full ACS (United States Census Bureau, 2018), allowing data users to create custom tables that are not available through pre-tabulated ACS data products. Using the PUMS files allowed comparison estimates to be calculated that correspond to the target population for the PATH Study.

[17] Missing values were imputed as part of the process to calibrate the Wave 4 Cohort weights to the 2016 ACS. This explains the slight differences observed between the PATH Study and 2016 ACS estimates for some characteristics in Table 5-3.

Table 5-3. Comparison of adult demographic and socio-economic estimates from PATH Study Wave 4 Cohort and ACS 2016

Wave 4 characteristic	PATH Study Wave 4 Cohort		ACS 2016
	Unweighted count	Weighted percentage [95% confidence interval]	Weighted percentage [95% confidence interval]
Sex			
Male	16,501	48.1% [47.6%, 48.6%]	48.1% [48.1%, 48.1%]
Female	17,109	51.9% [51.4%, 52.4%]	51.9% [51.9%, 51.9%]
Age group			
18-24	11,214	12.4% [12.0%, 12.7%]	12.4% [12.3%, 12.4]
25-44	11,335	34.0% [33.5%, 34.5%]	34.0% [34.0%, 34.0%]
45-64	7,985	34.0% [33.5%, 34.5%]	34.0% [34.0%, 34.1%]
65 and above	3,108	19.6% [19.2%, 20.0%]	19.6% [19.6%, 19.6%]
Race/ethnicity			
Hispanic	6,650	16.0% [15.6%, 16.3%]	15.7% [15.7%, 15.7%]
Non-Hispanic White alone	18,898	64.6% [64.1%, 65.1%]	64.3% [64.3%, 64.3%]
Non-Hispanic Black alone	4,952	11.3% [11.0%, 11.7%]	11.7% [11.7%, 11.7%]
Non-Hispanic other race or multiple races	2,616	8.2% [7.9%, 8.5%]	8.3% [8.3%, 8.3%]
Education			
Less than high school or GED	6,336	15.9% [15.5%, 16.3%]	16.0% [15.9%, 16.0%]
High school	8,114	23.8% [23.3%, 24.2%]	23.8% [23.7%, 23.9%]
Some college, no degree	11,899	31.0% [30.5%, 31.5%]	31.1% [31.0%, 31.2%]
Bachelor degree and above	7,147	29.3% [28.8%, 29.8%]	29.2% [29.1%, 29.3%]
Health insurance			
Yes	27,702	88.1% [87.5%, 88.6%]	91.9% [91.8%, 92.0%]
No	5,583	11.9% [11.4%, 12.5%]	8.1% [8.0%, 8.2%]

5.2.3 Comparison of Adult Cigarette-Use Estimates to Other National Studies

This section compares adult cigarette-smoking estimates based on the Wave 4 Cohort of the PATH Study to similar estimates based on data from the TUS-CPS 2014-2015, NHIS 2017, NHANES 2015-2016, and NSDUH 2016.

Table 5-4 presents estimates of the prevalence of current cigarette smoking[18] based on the Wave 4 Cohort, for the adult population as a whole and for subgroups. These estimates are accompanied by 95 percent confidence intervals. The point estimates for the PATH Study were calculated using the Wave 4 cross-sectional weights. The corresponding replicate weights were used to calculate variances and confidence intervals. Point estimates and 95 percent confidence intervals are reported for the other national studies as well.

Table 5-4 shows that the PATH Study estimates of adult current cigarette-smoking rates were similar to estimates from NHANES 2015-2016; these two studies had overlapping confidence intervals for almost all the estimates in the table. Estimates from Wave 4 of the PATH Study and NHANES 2015-2016 tended to be higher than those from TUS-CPS 2014-2015 and NHIS 2017, and lower than those from NSDUH 2016. Table 5-4 shows no evidence of nonresponse bias in the PATH Study with respect to current cigarette-smoking behavior among adults, in the sense that the PATH Study's estimates were all within the range of estimates from comparable surveys.

[18]For the PATH Study, following common practice for tobacco surveys, a current cigarette smoker is someone who (1) has smoked at least 100 cigarettes in his or her lifetime and (2) currently smokes cigarettes every day or some days. The questions used to define current cigarette smoking for each survey are provided in Appendix A.

Table 5-4. Comparison of adult cigarette-smoking estimates from PATH Study Wave 4 Cohort and other national studies*

Wave 4 outcome variable and subgroup	PATH Study Wave 4 Cohort sample size	Percentage from PATH Study Wave 4 Cohort [95% confidence interval]	Percentage from 2014-2015 TUS-CPS [95% confidence interval]	Percentage from 2017 NHIS [95% confidence interval]	Percentage from 2015-2016 NHANES [95% confidence interval]	Percentage from 2016 NSDUH, original definition[a] [95% confidence interval]	Percentage from 2016 NSDUH, modified definition[a] [95% confidence interval]
Current smoker, overall	33,608	17.7% [17.2%, 18.2%]	13.7% [13.5%, 13.9%]	14.0% [13.4%, 14.6%]	18.5% [16.6%, 20.4%]	20.7% [20.0%, 21.3%]	19.1% [18.5%, 19.7%]
Current smoker, male	16,477	19.9% [19.2%, 20.7%]	15.4% [15.1%, 15.7%]	15.8% [15.0%, 16.7%]	21.3% [18.7%, 24.2%]	22.8% [22.0%, 23.6%]	21.1% [20.3%, 21.8%]
Current smoker, female	17,097	15.7% [15.2%, 16.3%]	12.2% [11.9%, 12.5%]	12.2% [11.4%, 13.0%]	15.8% [14.4%, 17.4%]	18.7% [17.8%, 19.6%]	17.2% [16.4%, 18.1%]
Current smoker, age 18-24	11,204	14.5% [13.6%, 15.4%]	13.0% [12.2%, 13.8%]	10.4% [8.9%, 12.1%]	17.9% [14.1%, 22.4%]	N/A	N/A
Current smoker, age 25-44	11,325	23.0% [22.0%, 23.9%]	15.4% [15.0%, 15.7%]	16.1% [15.2%, 17.1%]	21.5% [19.2%, 24.0%]	N/A	N/A
Current smoker, age 45-64	7,976	19.3% [18.5%, 20.2%]	15.8% [15.4%, 16.1%]	16.5% [15.5%, 17.5%]	21.3% [18.8%, 24.0%]	N/A	N/A
Current smoker, age 65 and above	3,101	7.9% [7.0%, 8.9%]	7.5% [7.2%, 7.9%]	8.2% [7.4%, 9.0%]	8.6% [5.9%, 12.5%]	N/A	N/A
Current smoker, Hispanic	6,639	14.2% [13.2%, 15.2%]	9.1% [8.6%, 9.6%]	9.9% [8.7%, 11.2%]	15.7% [13.8%, 17.8%]	16.4% [15.0%, 18.0%]	13.7% [12.5%, 15.0%]
Current smoker, non-Hispanic White alone	18,889	18.8% [18.1%, 19.5%]	15.0% [14.8%, 15.3%]	15.2% [14.5%, 15.9%]	18.3% [15.5%, 21.5%]	22.1% [21.3%, 22.9%]	20.9% [20.1%, 21.7%]
Current smoker, non-Hispanic Black alone	4,943	20.8% [19.6%, 22.1%]	14.7% [14.0%, 15.4%]	14.9% [13.2%, 16.7%]	23.5% [21.1%, 26.0%]	21.6% [20.3%, 22.9%]	19.3% [18.0%, 20.6%]
Current smoker, non-Hispanic other race or multiple races	2,612	12.2% [11.0%, 13.5%]	10.7% [10.1%, 11.4%]	11.2% [9.6%, 13.0%]	18.1% [13.0%, 24.6%]	16.3% [14.2%, 18.5%]	14.6% [12.6%, 16.8%]
Current every-day smoker, overall	33,623	13.5% [13.1%, 14.0%]	10.6% [10.4%, 10.8%]	10.5% [9.9%, 11.0%]	14.2% [12.4%, 16.1%]	N/A	N/A
Current some-days smoker, overall	33,615	4.2% [4.0%, 4.4%]	3.1% [3.0%, 3.2%]	3.5% [3.2%, 3.8%]	4.3% [3.6%, 5.1%]	N/A	N/A

*Estimates from all the external sources were obtained using public use files for those surveys.

[a] NSDUH's definition of a current cigarette smoker is someone who has smoked part or all of a cigarette in the past 30 days, which is more expansive than the definition used in the other surveys. However, NSDUH contains questions on lifetime smoking and current smoking. The modified definition uses these questions to construct a measure of "current smoking" that is comparable to that of the other surveys (Ryan et al., 2012). The construction of this variable is described in Appendix A. Detailed age information was not available in the public use file for NSDUH 2016.

The disparities in Table 5-4 can be explained by a number of potential reasons. In addition to the varying degrees of sampling and measurement errors, the surveys differ in question order, context, mode of administration, and year of data collection. The TUS-CPS estimates of smoking prevalence are generally lower than estimates from the other surveys, which may be due to the proxy responses used in the TUS-CPS. The rotation group structure of the TUS-CPS may result in underestimates of smoking prevalence, as smokers are more likely to drop out over the course of the panel survey (Song, 2013). The PATH Study and NSDUH both use audio computer-assisted self-interview (ACASI) administration for the tobacco-use questions so that the interviewer does not see responses to the questions. In contrast, TUS-CPS, NHIS, and NHANES have direct questioning by an interviewer: NHIS and NHANES are conducted in person, and TUS-CPS is conducted in person and by telephone. The contexts and purposes of these surveys also differ: CPS is a general survey on unemployment, NHIS and NHANES are general health surveys, and NSDUH is a cross-sectional survey on substance use (including tobacco use) and health, including mental health. Unlike the cross-sectional prevalence surveys, the PATH Study uses a longitudinal cohort design to assess within-person changes and between-person differences in tobacco-use behaviors and health over time. Other differences among the questions used in the instruments of these different studies are outlined in Appendix A.

5.3 Youth Interview

5.3.1 Response Rates for Wave 4 Replenishment Sample

For the Wave 4 replenishment sample, the youth interview was based on completion of the household screener. Household screener response rates are reported in Table 5-1 in Section 5.1.1. The youth interview response rate (conditioning on completion of the household screener) was calculated as:

$$RR_Y = C_Y/(C_Y + N_Y)$$

where

RR_Y = Wave 4 replenishment sample youth interview response rate;

C_Y = number of Wave 4 youth interview complete cases; and

N_Y = number of Wave 4 youth interview nonrespondents.

Unweighted response rates were calculated based on actual case counts, while weighted response rates were based on the sums of the Wave 4 IPS weights (calculated as the product of the household IPS weight and the inverse of the within-household probability of selection).

Table 5-5 provides the youth interview response rates. In addition to the overall row, the table includes rows on age, sex, and race/ethnicity based on information collected through the household screener, as well as census region. Youth with missing values for a particular characteristic were excluded from the response rate calculation for that characteristic.

Table 5-5. Wave 4 youth interview response rates conditioning on household screener response (Wave 4 replenishment sample)

Characteristic reported by household screener respondent[a] or census region	C_Y: Youth interview completed (n)	N_Y: Youth interview nonresponse (n)	RR_Y: Unweighted response rate[b] (%)	RR_Y: Weighted response rate[b] (%)
Overall	3,739	1,574	70.4	70.6
Age group				
12-13	1,281	511	71.5	71.8
14-17	2,407	1,054	69.5	69.9
Sex				
Male	1,950	838	69.9	70.2
Female	1,787	736	70.8	71.0
Race/ethnicity				
Hispanic	1,147	393	74.5	74.7
Non-Hispanic White alone	1,720	820	67.7	67.8
Non-Hispanic Black alone	501	170	74.7	74.6
Non-Hispanic other race or multiple races	364	188	65.9	66.5
Census region				
Northeast	538	286	65.3	65.4
Midwest	803	359	69.1	68.9
South	1,438	566	71.8	72.0
West	960	363	72.6	73.0

[a] The sum of the counts in all the categories may not be equal to the count in the overall row due to missing values or unusual circumstances. For example, the sum of the counts across age categories is not equal to the count in the overall row because for a small proportion of cases, either the information for categorizing a case into the 12-13 or 14-17 category was unavailable, or the age reported by the household screener respondent was outside of the 12-17 range.

[b] The unweighted response rate = $C_Y/(C_Y+N_Y)$. The weighted response rate was calculated using a similar approach except that the Wave 4 IPS weights were used to obtain weighted counts.

The weighted overall response rate for the youth interview was 70.6 percent. The biggest subgroup difference occurred for race/ethnicity, with weighted response rates of 74.7 percent for Hispanic youth and 66.5 percent for non-Hispanic youth of other or multiple races. Differences among

subgroups on weighted response rates were 1.9 percentage points for age, 0.8 percentage points for sex, and 7.6 percentage points for census region.

5.3.2 Comparison of Wave 4 Cohort Demographic Estimates to the American Community Survey

Table 5-6 compares weighted estimates of youth demographic characteristics based on the Wave 4 Cohort youth with corresponding estimates based on the 2016 ACS PUMS. Estimates for the PATH Study were calculated using the Wave 4 Cohort cross-sectional weights and the ACS PUMS exclusions were the same as those described in Section 5.2.2 for the adult estimates.

Table 5-6. Comparison of youth demographic characteristics from PATH Study Wave 4 Cohort and ACS 2016

Wave 4 characteristic	PATH Study Wave 4 Cohort		ACS 2016
	Unweighted count	Weighted percentage [95% confidence interval]	Weighted percentage [95% confidence interval]
Sex			
Male	7,666	51.1% [50.3%, 51.9%]	51.2% [51.0%, 51.4%]
Female	7,075	48.9% [48.1%, 49.7%]	48.8% [48.6%, 49.0%]
Age group			
12-13	4,760	33.1% [33.1%, 33.1%]	33.1% [32.9%, 33.4%]
14-17	10,033	66.9% [66.9%, 66.9%]	66.9% [66.6%, 67.1%]
Race/ethnicity			
Hispanic	4,371	23.8% [23.1%, 24.5%]	23.5% [23.4%, 23.7%]
Non-Hispanic White alone	6,554	52.8% [52.0%, 53.6%]	52.9% [52.8%, 53.1%]
Non-Hispanic Black alone	1,912	13.4% [12.8%, 13.9%]	13.5% [13.4%, 13.7%]
Non-Hispanic other race or multiple races	1,387	10.0% [9.5%, 10.5%]	10.0% [9.9%, 10.1%]

The weighted percentages of sex, age, and race/ethnicity are nearly identical to the 1-year 2016 ACS estimates of those quantities. The Wave 4 Cohort cross-sectional weights were calibrated to match the 2016 ACS totals for these three characteristics, so this close match is the expected and desired

result. In Table 5-6, the youth with missing values for a particular characteristic were excluded from the Wave 4 Cohort estimates for that characteristic.[19]

Both the PATH Study and 2016 ACS estimates show a gender imbalance among 12- to 17-year-olds in the United States, with 51.2 percent being male and 48.8 percent being female. While this finding may not be intuitive, the United States Census Bureau has documented that the "sex ratio at birth in the United States has been around 105 males for every 100 females, however, since mortality at every age is generally higher for males, the sex ratio naturally declines with age" (https://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf).

5.3.3 Comparison of Youth Cigarette-Use Estimates to Other National Studies

Table 5-7 compares estimates for two common measures of youth cigarette-smoking based on the Wave 4 Cohort of the PATH Study to similar estimates based on data from NHANES 2015-2016, NSDUH 2016, and NYTS 2017.[20] The primary measure of cigarette smoking among youth is whether the youth has ever tried smoking a cigarette, even one or two puffs (see Appendix A). Another measure is current smoking, defined as having smoked at all in the past 30 days.

Estimates of both measures, accompanied by 95 percent confidence intervals, are shown for the youth population as a whole and for subgroups. The point estimates for the PATH Study were calculated using the Wave 4 cross-sectional weights. The corresponding replicate weights were used to calculate variances and confidence intervals. Point estimates and 95 percent confidence intervals are reported for the other national studies as well.

[19] Missing values were imputed as part of the process to calibrate the Wave 4 Cohort weights to the 2016 ACS. This explains the slight differences observed between the PATH Study and 2016 ACS estimates for some characteristics in Table 5-6.

[20] TUS-CPS does not interview persons younger than 18 about tobacco use.

Table 5-7. Comparison of youth cigarette-smoking estimates from PATH Study Wave 4 Cohort and other national studies*

Wave 4 outcome variable and subgroup	PATH Study Wave 4 Cohort sample size	Percentage from PATH Study Wave 4 Cohort [95% confidence interval]	Percentage from 2015-2016 NHANES [95% confidence interval]	Percentage from 2016 NSDUH [95% confidence interval]	Percentage from 2017 NYTS [95% confidence interval]
Ever tried cigarette smoking, even one or two puffs, overall	14,475	10.3% [9.7%, 10.8%]	12.9% [10.4%, 15.9%]	11.6% [10.8%, 12.5%]	15.9% [14.3%, 17.7%]
Ever tried smoking, male	7,496	11.0% [10.3%, 11.8%]	14.6% [10.1%, 20.6%]	12.3% [11.3%, 13.4%]	15.8% [14.1%, 17.6%]
Ever tried smoking, female	6,931	9.4% [8.6%, 10.3%]	11.2% [8.6%, 14.4%]	10.9% [9.9%, 12.1%]	15.8% [14.0%, 17.8%]
Ever tried smoking, age 12-13	4,752	3.0% [2.5%, 3.7%]	2.8% [1.1%, 6.7%]	3.1% [2.4%, 3.9%]	6.8% [5.6%, 8.2%]
Ever tried smoking, age 14-17	9,723	13.9% [13.2%, 14.8%]	17.5% [14.2%, 21.5%]	15.6% [14.5%, 16.7%]	20.4% [18.2%, 22.8%]
Have smoked in past 30 days, overall	14,770	3.1% [2.8%, 3.5%]	3.3% [2.0%, 5.2%]	3.5% [3.1%, 3.9%]	4.7% [4.0%, 5.4%]

* Estimates from all the external sources were obtained using public use files for those surveys.

The PATH Study point estimates tended to be lower than the estimates from the other studies. However, the confidence intervals overlapped between Wave 4 of the PATH Study, NHANES 2015-2016, and NSDUH 2016 for most estimates. As discussed in Section 5.2.3, some differences in the estimates may be attributable to the differences in question order and wording, mode of administration, and timeframes of the surveys.

Based on this analysis alone, it is unclear whether the PATH Study estimates of youth cigarette smoking are generally lower due to nonresponse bias in one or more of the estimates compared, or for other reasons. The PATH Study, NHANES, and NSDUH use ACASI for the questions about tobacco use by youth, and these are administered individually in a household or mobile examination center setting. The NYTS is a pencil-and-paper survey that is self-administered in the classroom. Currivan et al. (2004) found that even when telephone ACASI was used, estimates of youth smoking prevalence were lower for a telephone survey of youth smoking than for a school-based survey of the same population (see also Fowler and Stringfellow, 2001, for a discussion of higher smoking rates in school-based surveys). In addition, school-based surveys often include students who are older than 17, but age 17 is the upper age limit for youth in the PATH Study.

Finally, according to Monitoring the Future (Johnston et al., 2018), cigarette smoking among 8th to 12th graders dropped from 2015 to 2017 (the percentage of youth who had ever tried smoking dropped by 3.9 percentage points among 8th graders, 4.0 percentage points among 10th graders, and 4.5 percentage points among 12th graders). The lower percentages found in the PATH Study may reflect, in part, a continuation of this trend.

5.4 Adult Biospecimen Collections

5.4.1 Unweighted Response Rates

This section discusses unweighted response rates for the urine specimen collection and blood specimen collection among Wave 4 adult interview respondents who belong to the Wave 4 Cohort and were asked to provide a biospecimen.[21] Each response rate was calculated as a ratio, with the

[21] Weighted response rates are not provided because among the Wave 4 Cohort, the subset of Wave 1 adults asked to provide a urine specimen at Wave 4 does not represent a readily interpretable portion of the population. Similarly, among the Wave 4 Cohort, the first-time Adult Interview respondents were a mixture of adults of all ages from the Wave 4 replenishment sample and adults from the Wave 1 sample who were mostly age 18.

numerator being the number of Wave 4 Cohort adults who provided a biospecimen and the denominator being the number of Wave 4 Cohort adults from whom a biospecimen was requested.

Tables 5-8 and 5-9 show the unweighted response rates for urine collection and blood collection, respectively. In addition to the overall rows, the response rates are tabulated by Wave 4 socio-demographic characteristics, tobacco-use status, and recruitment wave (i.e., whether the person was sampled for the PATH Study at Wave 1 or Wave 4) for both urine and blood collection. For urine collection, response rates are also presented by whether Wave 4 was the first time the respondent had completed a PATH Study adult interview. Adults with missing values for a particular characteristic were excluded from the response rate calculation for that characteristic.

As shown in Table 5-8, the unweighted response rate for urine collection was 92.4 percent overall, with little variation by sex, race/ethnicity, education, and whether the adult respondent had health insurance. Current established tobacco users at Wave 4 were more likely to provide urine specimens (95.1 percent) than adults who were not current established users (90.4 percent), while adults ages 65 and above were the least likely to provide urine specimens (88.9 percent) among the age groups.

Adults from the Wave 4 replenishment sample and first-time adult interview respondents at Wave 4 were less likely to provide urine specimens; these are highly correlated characteristics because the majority of the first-time adult interview respondents were from the Wave 4 replenishment sample. This last finding is not surprising because almost all of the adults who had previously completed a PATH Study adult interview and were asked to provide a urine specimen at Wave 4, had provided urine in at least one previous wave. Despite the differences in response propensities across subgroups, the response rates for subgroups examined were at least 81 percent.

As shown in Table 5-9, the unweighted response rate for blood collection among first-time adult interview respondents (including adult respondents of any age selected as part of the replenishment sample) was 45.3 percent. Females, older adults, and non-Hispanic White alone adults were more likely to provide blood specimens than males (42.6 percent), younger adults, and other race/ethnicity groups. Similar to the results observed for the Wave 1 Cohort, blood collection response rates were lower for adults without health insurance and for current established tobacco users. The modest overall response rate for blood collection combined with moderate variation in response propensities by subgroups indicates some potential for nonresponse bias with respect to these data.

Table 5-8. Unweighted urine collection response rates among Wave 4 adult interview respondents selected to provide a urine specimen (Wave 4 Cohort)

Wave 4 characteristic[a] or recruitment wave	Wave 4 adult interview respondents selected to provide urine (n)	Urine providers among Wave 4 adult interview respondents (n)	Unweighted response rate for urine collection (%)
Overall	22,653	20,928	92.4
Sex			
Male	11,388	10,456	91.8
Female	11,247	10,456	93.0
Age group			
18-24	8,834	7,980	90.3
25-44	7,358	6,971	94.7
45-64	4,904	4,593	93.7
65 and above	1,557	1,384	88.9
Race/ethnicity			
Hispanic	4,609	4,277	92.8
Non-Hispanic White alone	12,433	11,443	92.0
Non-Hispanic Black alone	3,492	3,276	93.8
Non-Hispanic other race or multiple races	1,785	1,622	90.9
Education			
Less than high school or GED	4,724	4,429	93.8
High school	5,912	5,443	92.1
Some college, no degree	8,084	7,520	93.0
Bachelor degree and above	3,830	3,448	90.0
Health insurance			
Yes	18,287	16,874	92.3
No	4,101	3,837	93.6
Tobacco-use status[b]			
Current established user	9,688	9,210	95.1
Not current established user	12,876	11,640	90.4
First time completing adult interview			
Yes	7,947	6,561	82.6
No	14,706	14,367	97.7
Wave sampled for study			
Wave 1	16,588	16,013	96.5
Wave 4	6,065	4,915	81.0

[a] All the characteristics apply to the Wave 4 adult interview responses; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth or adult interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] A tobacco user is defined as someone who uses one or more of the following tobacco products covered by the Wave 4 adult interview: cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, and electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah). A 'current established user' of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly.

Table 5-9. Unweighted blood collection response rates among Wave 4 first-time adult interview respondents (Wave 4 Cohort)

Wave 4 characteristic[a] or recruitment wave	First-time adult interview respondents at Wave 4 (n)	Blood providers among first-time adult interview respondents at Wave 4 (n)	Unweighted response rate for blood collection (%)
Overall	7,947	3,602	45.3
Sex			
Male	4,104	1,750	42.6
Female	3,836	1,849	48.2
Age group			
18-24	4,531	1,967	43.4
25-44	1,689	746	44.2
45-64	1,242	645	51.9
65 and above	485	244	50.3
Race/ethnicity			
Hispanic	1,761	782	44.4
Non-Hispanic White alone	4,176	1,983	47.5
Non-Hispanic Black alone	1,165	468	40.2
Non-Hispanic other race or multiple races	700	307	43.9
Education			
Less than high school or GED	1,745	819	46.9
High school	2,438	1,069	43.8
Some college, no degree	2,448	1,123	45.9
Bachelor degree and above	1,266	575	45.4
Health insurance			
Yes	6,424	3,017	47.0
No	1,379	543	39.4
Tobacco-use status[b]			
Current established user	2,364	1,118	47.3
Not current established user	5,537	2,464	44.5
Wave sampled for study			
Wave 1	1,882	884	47.0
Wave 4	6,065	2,718	44.8

[a] All the characteristics apply to the Wave 4 adult interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth or adult interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] A tobacco user is defined as someone who uses one or more of the following tobacco products covered by the Wave 4 adult interview: cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, and electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah). A 'current established user' of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly.

5.4.2 Comparison Between First-Time Adult Interview Respondents Who Provided a Biospecimen at Wave 4 and All First-Time Adult Interview Respondents (Among the Wave 4 Cohort)

All first-time adult interview respondents at Wave 4 were asked to provide a urine specimen and a blood specimen. Table 5-10 compares the characteristics of Wave 4 first-time adult interview respondents who provided a urine specimen with all Wave 4 first-time adult interview respondents, among the Wave 4 Cohort. For this analysis, all adult respondents in the replenishment sample are considered first-time adult interview respondents at Wave 4. Table 5-11 presents a similar analysis for the collection of blood specimens. The Wave 4 cross-sectional weights (which account for Wave 4 interview nonresponse) were used for generating all the estimates.

Table 5-10 shows subgroup differences for some characteristics, however they were all below two percentage points. Among the urine providers, there is a slight underrepresentation of adults ages 65 and above, adults with at least a bachelor's degree, and adults with health insurance. Conversely, there is a slight overrepresentation of adults with less than high school or GED and current established users. According to Table 5-11, there were also differences between the characteristics of blood specimen providers and those asked to provide a blood specimen. Among the blood providers, males, younger adults (ages 18-24), and non-Hispanic adults whose race is not White alone were underrepresented and adults with health insurance and non-Hispanic White alone adults were overrepresented to some extent.

Note that this analysis applies to first-time adult interview respondents only, not to all Wave 4 adult interview respondents in the Wave 4 Cohort. Among this cohort, most first-time adult interview respondents at Wave 4 were from the Wave 4 replenishment sample, but some were sampled at Wave 1. The plan for laboratory analysis of Wave 4 adult biospecimens is unknown at this time, so the results in in Tables 5-10 and 5-11 may not correspond to those who will have Wave 4 biomarker data. A set of biospecimen weights may be created for analyzing the adult biomarker data, and those weights would attempt to correct for potential nonresponse bias associated with differential biospecimen collection response rates.

Table 5-10. Comparison of characteristics between Wave 4 first-time adult interview respondents who provided a urine specimen and all Wave 4 first-time adult interview respondents (Wave 4 Cohort)

Wave 4 characteristic[a]	First-time adult interview respondents at Wave 4		Urine providers among first-time adult interview respondents at Wave 4		Difference in weighted percentages [urine providers – first-time adult interview respondents at Wave 4] [95% confidence interval]
	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	
Sex					
Male	4,104	48.4% [47.3%, 49.5%]	3,368	48.1% [46.8%, 49.5%]	-0.2% [-1.4%, 0.9%]
Female	3,836	51.6% [50.5%, 52.7%]	3,188	51.9% [50.5%, 53.2%]	0.2% [-0.9%, 1.4%]
Age group					
18-24	4,531	30.8% [29.8%, 31.8%]	3,796	31.5% [30.4%, 32.6%]	0.7% [-0.1%, 1.6%]
25-44	1,689	28.1% [27.1%, 29.1%]	1,399	28.5% [27.1%, 29.9%]	0.4% [-0.8%, 1.6%]
45-64	1,242	25.9% [24.9%, 26.9%]	1,008	26.1% [25.1%, 27.2%]	0.2% [-0.7%, 1.2%]
65 and above	485	15.3% [14.5%, 16.1%]	358	13.9% [12.8%, 15.2%]	-1.4% [-2.4%, -0.4%]
Race/ethnicity					
Hispanic	1,761	18.2% [17.4%, 19.1%]	1,491	18.8% [17.9%, 19.8%]	0.6% [-0.2%, 1.4%]
Non-Hispanic White alone	4,176	57.5% [56.4%, 58.6%]	3,400	57.8% [56.6%, 59.0%]	0.3% [-0.7%, 1.3%]
Non-Hispanic Black alone	1,165	14.7% [13.9%, 15.5%]	990	14.7% [13.9%, 15.6%]	0.1% [-0.5%, 0.6%]
Non-Hispanic other race or multiple races	700	9.6% [9.0%, 10.3%]	556	8.7% [7.7%, 9.8%]	-1.0% [-2.1%, 0.2%]

Table 5-10. Comparison of characteristics between Wave 4 first-time adult interview respondents who provided a urine specimen and all Wave 4 first-time adult interview respondents (Wave 4 Cohort) (continued)

Wave 4 characteristic[a]	First-time adult interview respondents at Wave 4		Urine providers among first-time adult interview respondents at Wave 4		Difference in weighted percentages [urine providers – first-time adult interview respondents at Wave 4] [95% confidence interval]
	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	
Education					
Less than high school or GED	1,745	18.2% [17.4%, 19.1%]	1,503	19.3% [18.3%, 20.3%]	1.1% [0.5%, 1.8%]
High school	2,438	26.1% [25.1%, 27.0%]	2,045	26.1% [25.1%, 27.2%]	0.0% [-1.0%, 1.1%]
Some college, no degree	2,448	31.1% [30.1%, 32.1%]	2,013	31.5% [30.2%, 32.8%]	0.4% [-0.6%, 1.4%]
Bachelor degree and above	1,266	24.6% [23.7%, 25.6%]	964	23.1% [21.7%, 24.5%]	-1.6% [-2.9%, -0.2%]
Health insurance					
Yes	6,424	84.8% [83.0%, 86.4%]	5,297	83.7% [81.7%, 85.6%]	-1.1% [-1.6%, -0.5%]
No	1,379	15.2% [13.6%, 17.0%]	1,163	16.3% [14.4%, 18.3%]	1.1% [0.5%, 1.6%]
Tobacco-use status[b]					
Current established user	2,364	20.0% [19.0%, 21.1%]	2,005	21.2% [19.9%, 22.5%]	1.1% [0.5%, 1.7%]
Not current established user	5,537	80.0% [78.9%, 81.0%]	4,521	78.8% [77.5%, 80.1%]	-1.1% [-1.7%, -0.5%]

[a] All the characteristics apply to the Wave 4 adult interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth or adult interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] A tobacco user is defined as someone who uses one or more of the following tobacco products covered by the Wave 4 adult interview: cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, and electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah). A 'current established user' of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly.

Table 5-11. Comparison of characteristics between Wave 4 first-time adult interview respondents who provided a blood specimen and all Wave 4 first-time adult interview respondents (Wave 4 Cohort)

Wave 4 characteristic[a]	First-time adult interview respondents at Wave 4		Blood providers among first-time adult interview respondents at Wave 4		Difference in weighted percentages [blood providers – first-time adult interview respondents at Wave 4] [95% confidence interval]
	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	
Sex					
Male	4,104	48.4% [47.3%, 49.5%]	1,750	45.3% [42.9%, 47.6%]	-3.1% [-5.4%, -0.8%]
Female	3,836	51.6% [50.5%, 52.7%]	1,849	54.7% [52.4%, 57.1%]	3.1% [0.8%, 5.4%]
Age group					
18-24	4,531	30.8% [29.8%, 31.8%]	1,967	29.0% [27.6%, 30.5%]	-1.7% [-3.0%, -0.5%]
25-44	1,689	28.1% [27.1%, 29.1%]	746	26.9% [24.8%, 29.2%]	-1.2% [-3.1%, 0.8%]
45-64	1,242	25.9% [24.9%, 26.9%]	645	26.8% [24.9%, 28.9%]	0.9% [-1.0%, 2.9%]
65 and above	485	15.3% [14.5%, 16.1%]	244	17.2% [15.1%, 19.6%]	1.9% [-0.1%, 4.0%]
Race/ethnicity					
Hispanic	1,761	18.2% [17.4%, 19.1%]	782	17.1% [15.2%, 19.2%]	-1.1% [-3.0%, 0.9%]
Non-Hispanic White alone	4,176	57.5% [56.4%, 58.6%]	1,983	63.1% [60.6%, 65.5%]	5.6% [3.2%, 8.0%]
Non-Hispanic Black alone	1,165	14.7% [13.9%, 15.5%]	468	12.0% [10.8%, 13.4%]	-2.6% [-3.9%, -1.4%]
Non-Hispanic other race or multiple races	700	9.6% [9.0%, 10.3%]	307	7.7% [6.4%, 9.3%]	-1.9% [-3.4%, -0.4%]

Table 5-11. Comparison of characteristics between Wave 4 first-time adult interview respondents who provided a blood specimen and all Wave 4 first-time adult interview respondents (Wave 4 Cohort) (continued)

Wave 4 characteristic[a]	First-time adult interview respondents at Wave 4		Blood providers among first-time adult interview respondents at Wave 4		Difference in weighted percentages [blood providers – first-time adult interview respondents at Wave 4] [95% confidence interval]
	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	
Education					
Less than high school or GED	1,745	18.2% [17.4%, 19.1%]	819	19.0% [17.1%, 21.0%]	0.8% [-0.9%, 2.4%]
High school	2,438	26.1% [25.1%, 27.0%]	1,069	24.9% [22.9%, 27.0%]	-1.2% [-3.2%, 0.8%]
Some college, no degree	2,448	31.1% [30.1%, 32.1%]	1,123	31.6% [29.6%, 33.7%]	0.5% [-1.4%, 2.4%]
Bachelor degree and above	1,266	24.6% [23.7%, 25.6%]	575	24.5% [22.6%, 26.6%]	-0.1% [-2.1%, 1.9%]
Health insurance					
Yes	6,424	84.8% [83.0%, 86.4%]	3,017	86.8% [84.3%, 89.0%]	2.0% [0.5%, 3.5%]
No	1,379	15.2% [13.6%, 17.0%]	543	13.2% [11.0%, 15.7%]	-2.0% [-3.5%, -0.5%]
Tobacco-use status[b]					
Current established user	2,364	20.0% [19.0%, 21.1%]	1,118	20.7% [19.1%, 22.3%]	0.6% [-0.6%, 1.9%]
Not current established user	5,537	80.0% [78.9%, 81.0%]	2,464	79.3% [77.7%, 80.9%]	-0.6% [-1.9%, 0.6%]

[a] All the characteristics apply to the Wave 4 adult interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first Youth or adult interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] A tobacco user is defined as someone who uses one or more of the following tobacco products covered by the Wave 4 adult interview: cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, and electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah). A 'current established user' of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly.

5.5 Youth Urine Specimen Collection

5.5.1 Unweighted Response Rates

This section discusses unweighted response rates for the urine specimen collection among Wave 4 youth interview respondents who belong to the Wave 4 Cohort, all of whom were asked to provide urine specimen. Each response rate was calculated as a ratio, with the numerator being the number of Wave 4 Cohort youth who provided a urine specimen and the denominator being the number of Wave 4 youth interview respondents in the Wave 4 Cohort.

Table 5-12 shows the unweighted response rates for urine collection. In addition to the overall row, the response rates are tabulated by Wave 4 demographic characteristics and tobacco-use status. Youth with missing values for a particular characteristic were excluded from the response rate calculation for that characteristic.

The unweighted response rate for youth urine collection was 88.5 percent overall, with little variation by age. Males (91.0 percent) and youth who had ever used tobacco (93.8 percent) were more likely to provide a urine specimen than females (85.9 percent) and never users of tobacco (88.0 percent). The response rates varied by 6 percentage points across race/ethnicity groups, with non-Hispanic Black alone youth being the most likely to provide a specimen (93.4 percent). Youth from the Wave 4 replenishment sample were slightly less likely to provide urine specimens.

Table 5-12. Unweighted urine collection response rates among Wave 4 youth interview respondents (Wave 4 Cohort)

Wave 4 characteristic[a] or recruitment wave	Wave 4 youth interview respondents (n)	Urine providers among Wave 4 youth interview respondents (n)	Unweighted response rate for urine collection (%)
Overall	14,793	13,095	88.5
Sex			
Male	7,666	6,975	91.0
Female	7,075	6,076	85.9
Age group			
12-13	4,760	4,142	87.0
14-17	10,033	8,953	89.2
Race/ethnicity			
Hispanic	4,371	3,928	89.9
Non-Hispanic White alone	6,554	5,680	86.7
Non-Hispanic Black alone	1,912	1,773	92.7
Non-Hispanic other race or multiple races	1,387	1,218	87.8

Table 5-12. Unweighted urine collection response rates among Wave 4 youth interview respondents (Wave 4 Cohort) (continued)

Wave 4 characteristic[a] or recruitment wave	Wave 4 youth interview respondents (n)	Urine providers among Wave 4 youth interview respondents (n)	Unweighted response rate for urine collection (%)
Tobacco-use status[b]			
Ever user	1,506	1,413	93.8
Never user	12,969	11,407	88.0
Wave sampled for study			
Wave 1	11,054	9,890	89.5
Wave 4	3,739	3,205	85.7

[a] All the characteristics apply to the Wave 4 youth interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] An 'ever user' is someone who has ever used one or more of the tobacco products covered by the Wave 4 youth interview. A 'never user' is someone who has never used any of those tobacco products. Ever use of a tobacco product is defined as having ever used the product, even one or two times. The products covered by the Wave 4 youth interview are cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah), dissolvable tobacco, bidis, and kreteks.

5.5.2 Comparison Between Youth Interview Respondents Who Provided a Urine Specimen at Wave 4 and All Youth Interview Respondents (Among the Wave 4 Cohort)

All youth interview respondents in Wave 4 were asked to provide a urine specimen. Table 5-13 compares the characteristics of Wave 4 youth interview respondents who provided a urine specimen with all Wave 4 youth interview respondents, among the Wave 4 Cohort. The Wave 4 cross-sectional weights were used for generating all the estimates.

Consistent with the findings about subgroup response rates, Table 5-13 shows that compared to those youth respondents asked to a provide urine specimen, the urine providers slightly overrepresent males, older youth (ages 14-17), Hispanic and non-Hispanic Black alone youth, and youth who had ever used tobacco, and slightly underrepresent non-Hispanic White alone youth. Although the subgroup differences were statistically significant for these characteristics, all the differences were below two percentage points.

The plan for laboratory analysis of Wave 4 youth urine specimens is unknown at this time, so the results in Table 5-13 may not correspond to those who will have Wave 4 biomarker data. A set of biospecimen weights may be created for analyzing the youth biomarker data, and those weights would attempt to correct for potential nonresponse bias associated with differential urine collection response rates.

Table 5-13. Comparison of characteristics between Wave 4 youth interview respondents who provided a urine specimen and all Wave 4 youth interview respondents (Wave 4 Cohort)

Wave 4 characteristic[a]	Wave 4 youth interview respondents		Urine providers among Wave 4 youth interview respondents		Difference in weighted percentages [urine providers – Wave 4 youth interview respondents] [95% confidence interval]
	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using Wave 4 final weights [95% confidence interval]	
Sex					
Male	7,666	51.1% [50.3%, 51.9%]	6,975	52.7% [51.8%, 53.5%]	1.6% [1.2%, 1.9%]
Female	7,075	48.9% [48.1%, 49.7%]	6,076	47.3% [46.5%, 48.2%]	-1.6% [-1.9%, -1.2%]
Age group					
12-13	4,760	33.1% [33.1%, 33.1%]	4,142	32.4% [31.6%, 33.2%]	-0.7% [-1.0%, -0.4%]
14-17	10,033	66.9% [66.9%, 66.9%]	8,953	67.6% [66.8%, 68.4%]	0.7% [0.4%, 1.0%]
Race/ethnicity					
Hispanic	4,371	23.8% [23.1%, 24.5%]	3,928	24.3% [23.5%, 25.0%]	0.4% [0.2%, 0.7%]
Non-Hispanic White alone	6,554	52.8% [52.0%, 53.6%]	5,680	51.9% [51.0%, 52.7%]	-0.9% [-1.3%, -0.5%]
Non-Hispanic Black alone	1,912	13.4% [12.8%, 13.9%]	1,773	14.0% [13.4%, 14.6%]	0.7% [0.4%, 0.9%]
Non-Hispanic other race or multiple races	1,387	10.0% [9.5%, 10.5%]	1,218	9.8% [9.3%, 10.4%]	-0.1% [-0.4%, 0.1%]
Tobacco-use status[b]					
Ever user	1,506	10.3% [9.7%, 10.8%]	1,413	10.9% [10.3%, 11.6%]	0.7% [0.5%, 0.8%]
Never user	12,969	89.7% [89.2%, 90.3%]	11,407	89.1% [88.4%, 89.7%]	-0.7% [-0.8%, -0.5%]

[a] All the characteristics apply to the Wave 4 youth interview; however, respondents are asked to self-report their sex and race/ethnicity only once, during their first youth interview. For each characteristic, the sum of the counts in all the categories may not be equal to the count in the overall row due to missing values.

[b] An 'ever user' is someone who has ever used one or more of the tobacco products covered by the Wave 4 youth interview. A 'never user' is someone who has never used any of those tobacco products. Ever use of a tobacco product is defined as having ever used the product, even one or two times. The products covered by the Wave 4 youth interview are cigarettes, traditional cigars, cigarillos, filtered cigars, pipes, smokeless tobacco, snus, hookah, electronic nicotine delivery systems or ENDS (including e-cigarettes, e-cigars, e-pipes, and e-hookah), dissolvable tobacco, bidis, and kreteks.

6. Summary of Findings

This chapter summarizes the PATH Study's Wave 4 nonresponse bias analysis findings separately for the Wave 1 Cohort and the Wave 4 Cohort. Each of the following components of the analysis is reviewed:

- Wave 4 screening and interview response rates.
- The national representativeness of Wave 4 respondents and statistical weighting adjustments to reduce potential nonresponse bias.
- Nonresponse to the Wave 4 biospecimen collections among Wave 4 adult and youth interview respondents.

6.1 Wave 1 Cohort

6.1.1 Adult Interview

The Wave 1 Cohort unweighted response rate was similar to the weighted response rate for the Wave 4 adult interview (conditioning on Wave 1 response) overall and by subgroups, although there were differences in weighted response rates across subgroups.

- The weighted overall response rate was 73.5 percent.
- The weighted response rate was lower among males than among females; this pattern is consistent with most household surveys.
- Non-Hispanic adults of other or multiple races and non-Hispanic White alone adults had moderately lower weighted response rates than Hispanic adults and non-Hispanic Black alone adults.

However, differences in response rates do not necessarily indicate nonresponse bias in the Wave 4 adult estimates for the Wave 1 Cohort.

As a result of comparing select demographic and socio-economic characteristics between the Wave 4 adult interview respondents and those eligible for the Wave 4 adult interview, the notable underrepresentation due to panel attrition among the Wave 1 Cohort was for the male population,

which tends to exhibit lower response propensity than the female population in most household surveys.

After Wave 4 longitudinal weighting adjustments, the differences between the Wave 1 Cohort respondents at Wave 4 and those eligible for the adult interview were negligible with the exception of the Wave 1 "ever tobacco use" measure for non-Hispanic Black alone adults who were youth at Wave 1. These results reflect the use of both demographic variables and tobacco-use measures from Wave 1 to calibrate the Wave 4 single-wave weights.

6.1.2 Youth Interview

The Wave 1 Cohort unweighted response rate was similar to the weighted response rate for the Wave 4 youth interview (conditioning on Wave 1 response) overall and by subgroups, although there were differences in weighted response rates across subgroups.

- The weighted overall response rate was 79.5 percent.
- Moderate differences (i.e., of no more than 5 percent) were found by Wave 1 age and race/ethnicity.
- The weighted response rate for those under age 12 was noticeably lower than for those ages 12-13 or 14-17 at Wave 1. This is likely because the persons in the youngest age group were shadow youth at Wave 1, so for some of them the Wave 4 youth interview was the first PATH Study interview they were asked to complete. (A similar response rate pattern was observed in Waves 2 and 3.)
- The weighted response rate among Wave 1 youth was lower for the Wave 1 "ever tobacco user" group than for the Wave 1 "never user" group.

These differences in response rates do not necessarily indicate nonresponse bias in the Wave 4 youth estimates for the Wave 1 Cohort as variation in response rates by subgroups is to be expected in large-scale data collection efforts.

As a result of comparing select demographic and tobacco-use characteristics between the Wave 4 youth interview respondents and those eligible for the Wave 4 youth interview, the notable underrepresentation due to panel attrition among the Wave 1 Cohort was for the Wave 1 "under 12" age group, i.e., Wave 1 shadow youth.

After Wave 4 longitudinal weighting adjustments, the underrepresentation of Wave 1 shadow youth improved but was not entirely eliminated. Estimates of Wave 1 "ever tobacco use" among those who were youth at Wave 1 remained slightly lower for the Wave 4 respondents than for those eligible for the youth interview overall, and for Wave 1 youth ages 14-17 and non-Hispanic White alone youth.

6.1.3 Adult Biospecimen Collections

Some Wave 4 adult interview respondents in the Wave 1 Cohort were asked to provide biospecimens. A subsample of the respondents in the Wave 1 Cohort were asked to provide a urine specimen and all first-time adult interview respondents were asked to provide a blood specimen. Unweighted response rates were calculated conditioning on Wave 4 interview completion.

- The unweighted overall response rate for urine collection was 96.5 percent.
- The unweighted urine collection response rate was noticeably lower for those completing the adult interview for the first time (87.4 percent).
- The unweighted overall response rate for blood collection was 46.8 percent.
- Females, Hispanic adults, and adults with health insurance were more likely to provide blood specimens than males, non-Hispanic adults, and adults without health insurance, but blood collection response rates varied little by tobacco-use status.

As a result of comparing select socio-demographic characteristics between the first-time adult interview respondents at Wave 4 who provided a biospecimen and those who were asked to do so, small differences were observed for the Wave 1 Cohort by race/ethnicity for both urine and blood collections, and males were underrepresented among blood providers. However, these estimates are based on the Wave 4 single-wave interview weights. A set of biospecimen weights may be created for analyzing the adult biomarker data; such weights would attempt to correct for potential nonresponse bias associated with differential biospecimen collection response rates.

6.1.4 Youth Urine Specimen Collection

All Wave 4 youth interview respondents were asked to provide urine specimens and unweighted response rates were calculated conditioning on Wave 4 interview completion. Among those belonging to the Wave 1 Cohort:

- The unweighted overall response rate for urine collection was 89.4 percent.
- Males and youth who had ever used tobacco were more likely to provide urine specimens than females and never users.

As a result of comparing select demographic characteristics between the Wave 4 youth interview respondents who provided a urine specimen and those who were asked to do so, small differences were observed for most of the Wave 1 Cohort subgroups. However, as is the case for adults, a set of biospecimen weights may be created for analyzing the youth biomarker data.

6.2 Wave 4 Cohort

6.2.1 Household Screener

Among the households screened for selection of Wave 4 replenishment sample adults and youth, the unweighted response rate was similar to the weighted response rate overall and by subgroups. However, there were differences in weighted response rates across subgroups.

- The weighted overall response rate was 52.8 percent.
- The weighted response rate for households in "high" education neighborhoods was 9.9 percentage points lower than in "low" education neighborhoods.
- The weighted response rate for households in "high" poverty neighborhoods was 7.5 percentage points higher than in "low" poverty neighborhoods.
- The weighted response rate for households in the Midwest was noticeably higher than for the other census regions.

6.2.2 Adult Interview

This response rate summary applies only to the Wave 4 Cohort adults selected as part of the Wave 4 replenishment sample. For these adults, the unweighted response rate was similar to the weighted

response rate for the Wave 4 adult interview (conditioning on household response) overall and by subgroups, although there were differences in weighted response rates across subgroups.

- The weighted overall response rate was 68.0 percent.
- The weighted response rate was noticeably lower among adults living in the Northeast census region and non-Hispanic adults of other or multiple races.
- Males, younger adults (ages 18-24), and those sampled as a nonuser of tobacco had moderately higher weighted response rates than females, older adults, and those sampled as a tobacco user.

Weighted demographic estimates based on all Wave 4 Cohort adults were nearly identical to corresponding estimates based on the 2016 ACS, consistent with the calibration of the Wave 4 cross sectional weights. However, adults with health insurance were slightly underrepresented by the Wave 4 Cohort estimate.

There is no evidence of nonresponse bias in the Wave 4 Cohort with respect to current cigarette-smoking behavior among adults, in the sense that the PATH Study's estimates were all within the range of estimates from other national studies. In general, the estimates of adult current cigarette-smoking rates based on the Wave 4 Cohort were similar to estimates from NHANES 2015-2016, higher than estimates from TUS-CPS 2014-2015 and NHIS 2016, and lower than estimates from NSDUH 2016.

6.2.3 Youth Interview

This response rate summary applies only to the Wave 4 Cohort youth selected as part of the Wave 4 replenishment sample. For these youth, the unweighted response rate was similar to the weighted response rate for the Wave 4 youth interview (conditioning on household response) overall and by subgroups, although there were differences in weighted response rates across subgroups.

- The weighted overall response rate was 70.6 percent.
- The weighted response rates varied by 8.2 percentage points for race/ethnicity and 7.6 percentage points for census region. Response rates were higher among Hispanic and non-Hispanic Black alone youth and youth living in the South or West census regions.

Weighted demographic estimates based on all Wave 4 Cohort youth were nearly identical to corresponding estimates based on the 2016 ACS, consistent with the calibration of the Wave 4 cross sectional weights.

The PATH Study estimates of youth cigarette-smoking rates tended to be lower than the estimates from the other studies. However, the confidence intervals overlapped between the Wave 4 Cohort of the PATH Study, NHANES 2015-2016, and NSDUH 2016 for most estimates.

6.2.4 Adult Biospecimen Collections

For the Wave 4 Cohort, all of the adult interview respondents from the replenishment sample and a subsample of the adult interview respondents from the Wave 1 sample were asked to provide a urine specimen. Only first-time adult interview respondents (including adult respondents of all ages from the Wave 4 replenishment sample) were asked to provide blood specimens. Unweighted response rates were calculated conditioning on Wave 4 interview completion.

- The unweighted overall response rate for urine collection was 92.4 percent.
- The unweighted urine collection response rate was noticeably lower for those completing the adult interview for the first time (82.6 percent). This is not surprising because almost all of the adults who had previously completed a PATH Study adult interview and were asked to provide a urine specimen at Wave 4, had provided urine at least one previous wave.
- The unweighted overall response rate for blood collection was 45.3 percent.
- The unweighted response rates varied moderately (i.e., by no more than 9 percent) by sex, age, race/ethnicity, and health insurance. Response rates were higher among females, older adults, non-Hispanic White alone adults, and adults with health insurance.

For blood collection, males, non-Hispanic White alone persons, 18- to 24-year-olds, and those with health insurance were less likely to provide a specimen.

As a result of comparing select socio-demographic characteristics between the first-time adult interview respondents at Wave 4 who provided a biospecimen and those who were asked to do so, small differences (below 2 percentage points for urine collection and 6 percentage points for blood collection) were observed for several subgroups in the Wave 4 Cohort. However, these estimates are

based on the Wave 4 cross-sectional interview weights. A set of biospecimen weights may be created for analyzing the adult biomarker data; such weights would attempt to correct for potential nonresponse bias associated with differential biospecimen collection response rates.

6.2.5 Youth Urine Specimen Collection

All Wave 4 youth interview respondents were asked to provide urine specimens and unweighted response rates were calculated conditioning on Wave 4 interview completion. Among those belonging to the Wave 4 Cohort:

- The unweighted overall response rate for urine collection was 88.5 percent.
- Males, youth who had ever used tobacco, and youth sampled at Wave 1 were more likely to provide urine specimens than females, never users, and youth from the Wave 4 replenishment sample.
- Moderate differences (i.e., of no more than 6 percent) were found by race/ethnicity.

As a result of comparing select demographic characteristics between the Wave 4 youth interview respondents who provided a urine specimen and those who were asked to do so, small differences were observed for most of the Wave 4 Cohort subgroups. However, as is the case for adults, a set of biospecimen weights may be created for analyzing the youth biomarker data.

6.3 General Conclusions

Assuming that the demographic, socio-economic, and tobacco-use characteristics examined in this report are correlated with key tobacco and health related outcome measures, these results raise no serious concern about potential nonresponse bias in estimates from Wave 4 of the PATH Study. However, they hint at the challenges faced by cohort studies as attrition accumulates and confirm the timeliness of the introduction of a replenishment sample to the study.

The findings for the Wave 1 Cohort suggest that the longitudinal weighting adjustments satisfactorily address the potential bias due to attrition. Similarly, the cross-sectional weights developed for the newly-established Wave 4 Cohort produce estimates that closely align with those from external data sources.

References

Bose, J., and West, J. (2002). Examining additional nonresponse bias introduced through attrition. *Proceedings of the Survey Research Methods Section, American Statistical Association,* 278-283.

Brick, J.M., Lê, T., and West, J. (2003). Dealing with movers in a longitudinal study of children. In *Statistics Canada Symposium-challenges in survey taking for the next decade.*

Brownstein, N., Kalsbeek, W.D., Tabor, J., Entzel, P., Daza, E., and Harris, K.M. (2009). *Non-Response in Wave IV of the National Longitudinal Study of Adolescent Health.* http://www.cpc.unc.edu/projects/addhealth/data/guides/W4_nonresponse.pdf.

Center for Behavioral Health Statistics and Quality (CBHSQ). *National Survey on Drug Use and Health,* Rockville, MD: Substance Abuse and Mental Health Services Administration, 2016, http://datafiles.samhsa.gov/study-dataset/national-survey-drug-use-and-health-2016-nsduh-2016-ds0001-nid17185.

Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). *National Health and Nutrition Examination Survey Data.* Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2015-2016 https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2015.

Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). *National Health Interview Survey Data.* Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2017, http://www.cdc.gov/nchs/nhis/nhis_2017_data_release.htm.

Centers for Disease Control and Prevention (CDC). *National Youth Tobacco Survey Data.* Atlanta, GA: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2017, https://www.cdc.gov/tobacco/data_statistics/surveys/nyts/data/index.html.

Cunradi, C. B., Moore, R., Killoran, M., and Ames, G. (2005). Survey nonresponse bias among young adults: the role of alcohol, tobacco, and drugs. *Substance Use & Misuse*, 40, 171-185.

Currivan, D.B., Nyman, A.L., Turner, C.F., and Biener, L. (2004). Does telephone audio computer-assisted self-interviewing improve the accuracy of prevalence estimates of youth smoking? Evidence from the UMass Tobacco Study. *Public Opinion Quarterly*, 68, 542-564.

Fowler, F.J., and Stringfellow, V.L. (2001). Learning from experience: Estimating teen use of alcohol, cigarettes, and marijuana from three survey protocols. *Journal of Drug Issues*, 31, 643-664.

Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70, 646-675.

Groves, R.M. and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. New York, NY: Wiley.

Hanushek, E.A., and Jackson, J.E. (1977). *Statistical Methods for Social Scientists*. New York, NY: Academic Press.

Heeringa, S.G., West, B.T., and Berglund, P.A. (2010). *Applied Survey Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC.

Johnston, L.D., Miech, R.A., O'Malley, P.M., Bachman, J.G., Schulenberg, J.E., and Patrick, M.E. (2018). *Monitoring the Future national survey results on drug use: 1975-2017: Overview, key findings on adolescent drug use*. Ann Arbor: Institute for Social Research, The University of Michigan. Available at: http://www.monitoringthefuture.org/pubs/monographs/mtf-overview2017.pdf.

Javitz, H., and Wagner, M. (2005). *Analysis of potential bias in the Wave 1 and Wave 2 respondents to the National Longitudinal Transition Study-2 (NLTS-2)*. Menlo Park, CA: SRI International, http://www.nlts2.org/studymeth/nlts2_analysis_bias_respondents.pdf.

Lundström, S., and Särndal, C-E (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, 15 (2), 305-327.

Office of Management and Budget, and Office of Management and Budget. "Standards and guidelines for statistical surveys." (2006). Available at: https://unstats.un.org/unsd/dnss/docs-nqaf/USA_standards_stat_surveys.pdf

Ryan, H., Trosclair, A., and Gfroerer, J. (2012). Adult current smoking: Differences in definitions and prevalence estimates—NHIS and NSDUH, 2008. *Journal of Environmental and Public Health,* online.

Särndal, C.-E., and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. Hoboken, NJ: Wiley.

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33, 99-119.

SAS Institute, Inc. (2013). *SAS® 9.4 Help and Documentation*. 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, 182-186.

Song, Y. (2013). *Rotation group bias in smoking prevalence estimates using TUS-CPS*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington DC, paper available at https://nces.ed.gov/FCSM/pdf/I3_Song_2013FCSM.pdf, and slides available at http://www.copafs.org/UserFiles/file/fcsm/I3_Song_2013FCSM.pdf.

Stoop, I.A.L. (2005). *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*. The Hague: Social and Cultural Planning Office.

United Nations. (2005). *Designing Household Survey Samples: Practical Guidelines.* United Nations Publication ST/ESA/STAT/SER.F/98, New York: United Nations. Available at http://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf.

United States Department of Commerce, Census Bureau (2016). National Cancer Institute and Food and Drug Administration co-sponsored Tobacco Use Supplement to the Current Population Survey. 2014-15. https://cancercontrol.cancer.gov/brp/tcrb/tus-cps/, https://thedataweb.rm.census.gov/ftp/cps_ftp.html#cpssupps.

Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22, 209-212.

Young, A.F., Powers, J.R., and Bell, S.L. (2006). Attrition in longitudinal studies: Who do you lose? *Australian and New Zealand Journal of Public Health,* 30, 353-361.

Appendix A
Cigarette-Smoking Questions in the PATH Study and Other Surveys

Table A-1 lists the questions used to ask about current smoking status of adults in the PATH Study and in the surveys used for comparison and describes the populations included in the estimates from those surveys.

Note that although the questions used to define current cigarette smoking are similar among the surveys, small differences could have an effect on the answers given. In the PATH Study, the question used to establish whether a respondent has smoked at least 100 cigarettes in his or her lifetime has closed response categories:

"How many cigarettes have you smoked in your entire life? A pack usually has 20 cigarettes in it."

For adults:

1. 1 or more puffs but never a whole cigarette;
2. 1 to 10 cigarettes (about ½ pack total);
3. 11 to 20 cigarettes (about ½ pack to 1 pack);
4. 21 to 50 cigarettes (more than 1 pack but less than 3 packs);
5. 51 to 99 (more than 2½ packs but less than 5 packs); and
6. 100 or more cigarettes (5 packs or more).

For youth:

1. 1 or more puffs but never a whole cigarette;
2. 1 cigarette;
3. 2 to 10 cigarettes (about ½ pack total);
4. 11 to 20 cigarettes (about ½ pack to 1 pack);
5. 21 to 50 cigarettes (more than 1 pack but less than 3 packs);

6. 51 to 99 (more than 2½ packs but less than 5 packs); and

7. 100 or more cigarettes (5 packs or more).

In TUS-CPS, NHIS, and NHANES, however, the question "Have you smoked at least 100 cigarettes in your entire life?" calls for a yes/no response.

Positioning of the questions also differs among the surveys. In the PATH Study, the cigarette smoking questions are near the beginning of the adult questionnaire, and the respondent knows that the questionnaire is about tobacco-use behaviors. In TUS-CPS, the smoking questions are near the beginning of the adult questionnaire on tobacco, but the survey is administered as part of the CPS. In NHIS, the smoking questions follow a long series of questions on health problems (breathing problems, diabetes, hernias, hemorrhoids, etc.). These differences among the surveys may be associated with differences in responses to the cigarette-smoking questions.

Table A-2 lists the questions used to define youth cigarette smoking in the PATH Study, NHANES, NSDUH, and NYTS.

Table A-1. Questions used to define adult current cigarette smoking in the PATH Study, TUS-CPS, NHIS, NHANES, and NSDUH

PATH Study	TUS-CPS*	NHIS	NHANES	NSDUH (original definition)	NSDUH (modified definition)**
Question to define current smoking (answers defining current smoking given in parentheses)					
["Have you ever smoked a cigarette, even one or two puffs?" (Wave 1 adult and Wave 4 replenishment sample adult question, yes) and "Have you ever tried cigarette smoking, even one or two puffs?" (Wave 1–3 baseline youth question, yes) and "In the past 30 days, have you smoked a cigarette, even one or two puffs?" (Wave 2–4 adult question, yes) and "In the past 12 months, have you smoked a cigarette, even one or two puffs" (Wave 2–4 adult and Wave 2–3 youth question, yes)] and "Do you now smoke cigarettes every day, some days, or not at all?" (every day or some days) and "How many cigarettes have you smoked in your entire life? A pack usually has 20 cigarettes in it." (Wave 1–4 adult and Wave 1–3 youth question, 100 or more cigarettes (5 packs or more))	"Have you smoked at least 100 cigarettes in your entire life?" (yes) and "Do you now smoke cigarettes every day, some days, or not at all?" (every day or some days)	"Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" (yes) and "Do you NOW smoke cigarettes every day, some days or not at all?" (every day or some days)	"{Have you/Has SP} smoked at least 100 cigarettes in {your/his/her} entire life?" (yes) and "{Do you/Does SP} now smoke cigarettes every day, some days, or not at all?" (every day or some days)	"Have you ever smoked part or all of a cigarette?" (yes) and "During the past 30 days, have you smoked part or all of a cigarette?" and "How long has it been since you last smoked part or all of a cigarette?" (within the past 30 days)	"Have you ever smoked part or all of a cigarette?" (yes) and "During the past 30 days, have you smoked part or all of a cigarette?" and "How long has it been since you last smoked part or all of a cigarette?" (within the past 30 days) and "Have you smoked at least 100 cigarettes in your entire life?" (yes)
Age range included in estimates					
18+	18+	18+	18+	18+	18+

Table A-1. Questions used to define adult current cigarette smoking in the PATH Study, TUS-CPS, NHIS, NHANES, and NSDUH (continued)

PATH Study	TUS-CPS*	NHIS	NHANES	NSDUH (original definition)	NSDUH (modified definition)**
Exclusions from population					
The Wave 4 Cohort target population included only the U.S. civilian, noninstitutionalized population at Wave 4.	Includes only the U.S. civilian, non-institutionalized population.	Includes only the civilian noninstitutionalized population residing in the U.S. at the time of the interview. Several segments of the population are excluded, such as: persons in long-term care institutions; persons on active duty with the Armed Forces; persons in correctional facilities; and U.S. nationals living in foreign countries.	Includes only the U.S. civilian, non-institutionalized population.	Includes only the U.S. civilian, non-institutionalized population. Excludes homeless persons who do not use shelters, military personnel on active duty, and residents of institutional group quarters, such as jails and hospitals.	Includes only the U.S. civilian, non-institutionalized population. Excludes homeless persons who do not use shelters, military personnel on active duty, and residents of institutional group quarters, such as jails and hospitals.
Proxy responses allowed					
No	Yes	Yes, for individuals physically or mentally incapable of responding.	Yes, when there is a serious physical or mental condition, a proxy respondent may be used to conduct the interview.	No	No

*Proxies are allowed if fourth callback or the person will not return before closeout. See https://cancercontrol.cancer.gov/brp/tcrb/tus-cps/surveys/A1_TUS-CPS-Questionnaire.pdf, p2.

**The modified definition is given in Ryan et al. (2012).

Table A-2. Questions used to define youth cigarette smoking in the PATH Study, NHANES, NSDUH, and NYTS

PATH Study	NHANES	NSDUH	NYTS
Question to define ever tried cigarette smoking (answers defining ever tried cigarette smoking given in parentheses)			
Have you ever tried cigarette smoking, even one or two puffs?" (Wave 1 youth and Wave 2–4 new baseline youth question, yes) and "In the past 12 months, have you smoked a cigarette, even one or two puffs?" (Wave 2–4 question for continuing youth, yes)	"About how many cigarettes have you smoked in your entire life?" (1 or more puffs to 100 or more cigarettes)	Have you ever smoked part or all of a cigarette?" (yes)	"Have you ever tried cigarette smoking, even one or two puffs?" (yes)
Questions for determining whether have smoked in past 30 days			
["Have you ever tried cigarette smoking, even one or two puffs?" (Wave 1 youth and Wave 2–4 new baseline youth question, yes) and "In the past 12 months, have you smoked a cigarette, even one or two puffs?" (Wave 2–4 question for continuing youth, yes)] and "When was the last time you smoked a cigarette, even one or two puffs?" (earlier today, not today but sometime in the past 7 days, not in the past 7 days but sometime in the past 30 days)	"On how many of the past 30 days did {you/SP} smoke a cigarette?" (1-30)	"Have you ever smoked part or all of a cigarette?" (yes) and ["During the past 30 days, have you smoked part or all of a cigarette?" (yes) and "During the past 30 days, that is since [DATEFILL], on how many days did you smoke part or all of a cigarette?" (1-30) and "What is your best estimate of the number of days you smoked part or all of a cigarette during the past 30 days?" (1 or 2 days to all 30 days)]	"During the past 30 days, on how many days did you smoke cigarettes?" (1-30)
Age range included in estimates			
12-17	12-17	12-17	12-17

Table A-2. Questions used to define youth cigarette smoking in the PATH Study, NHANES, NSDUH, and NYTS (continued)

PATH Study	NHANES	NSDUH	NYTS
Exclusions from population			
The Wave 4 Cohort target population included only the U.S. civilian, noninstitutionalized population at Wave 4.	Includes only the U.S. civilian, noninstitutionalized population.	Includes only the U.S. civilian, noninstitutionalized population. Excludes homeless persons who do not use shelters, military personnel on active duty, and residents of institutional group quarters, such as jails and hospitals.	Only includes youth who are public and private school students enrolled in regular middle schools and high schools in grades 6 through 12 in the 50 U.S. States and the District of Columbia. Alternative schools, special education schools, Department of Defense operated schools, Bureau of Indian Affairs schools, vocational schools that serve only pull-out populations, and students enrolled in regular schools unable to complete the questionnaire without special assistance, are excluded.

Table A-2. Questions used to define youth cigarette smoking in the PATH Study, NHANES, NSDUH, and NYTS (continued)

PATH Study	NHANES	NSDUH	NYTS
Other comments			
	Youth with missing values for the response to the question about number of lifetime cigarettes smoked were excluded from the estimates of ever tried cigarette smoking. Youth with missing values for the response to the question about number of cigarettes smoked in the past 30 days were excluded from the estimates of past 30 day cigarette use unless the value was missing because the youth had never smoked a cigarette in his/her lifetime. Youth who had never smoked were treated as having smoked zero cigarettes in the past 30 days.		Self-administered survey in classroom.

Appendix B
Wave 4 Replenishment Sample: Household Screener and Shadow Youth Response Rates

B.1 Household Screener Response Rates for Wave 4 Replenishment Sample - Combining the Adult, Youth, and Shadow Youth (AYS) and the Shadow Youth Only (SO) Samples

The Wave 4 replenishment sample was comprised of two separate address-level samples that were screened for different purposes (see Section 2.3). The purpose of the Adult, Youth, and Shadow Youth (AYS) sample was to identify and interview enough adults and youth to replace those from the Wave 1 Cohort lost to attrition and to maintain desired levels of precision for key outcomes of interest. In addition, if the household contained a 10- or 11-year-old child and the child was selected for the PATH Study, a parent (or guardian) was asked to consent to the selected child joining the study as a shadow youth member for interviewing at a later wave after turning 12.

All PATH Study participants recruited at Wave 1 were expected to be age 12 or older by the time of Wave 4. Thus, a much larger number of sampled addresses were required to recruit 10- and 11-year-olds at Wave 4 than new participants 12 years old and older. To achieve the desired number of 10- and 11-year-old participants, a Shadow Youth Only (SO) sample of addresses was selected in addition to the AYS sample. The addresses selected for the SO sample were screened only for 10- and 11-year-olds.

Section 5.1 presented household screener response rates for the Wave 4 AYS replenishment sample. This appendix provides household screener response rates for the entire Wave 4 replenishment sample, i.e., the AYS and SO samples combined. These response rates were calculated using the formula described in Section 5.1.1. Unweighted counts and weighted counts based on the Wave 4 replenishment sample household IPS weights were computed for each response status category, for the unweighted response rates and weighted response rates, respectively.

Table B-1 provides overall response rates for the household screener and response rates by subgroups of sampled households. Subgroups were defined by the characteristics of the census

block groups in which the sampled addresses were located, using information from the 5-year (2012 to 2016) ACS.[22] The "high" and "low" subgroup categories were defined relative to the nationwide percentage of persons having the characteristic: block groups with percentages below the national average for the characteristic were classified as low, and those with percentages above the national average were classified as high.

Table B-1. Wave 4 household screener response rates for the entire replenishment sample

Neighborhood characteristic[b] or census region	C_{HH}: Completed (n)	N_{HH}: Nonresponse known to be eligible (n)	$e_{HH} \times U_{HH}$: Nonresponse with unknown eligibility estimated to be eligible (n)	RR_{HH}: Unweighted response rate[a] (%)	RR_{HH}: Weighted response rate[a] (%)
Overall	79,540	66,437	8,964	51.3	51.2
Education (% of persons ages 25+ with bachelor's degree)					
High (> 30.3%)	33,368	29,739	5,466	48.7	48.4
Low (<= 30.3%)	46,172	36,698	3,498	53.5	53.4
Race (% Black alone)					
High (> 12.6%)	19,048	16,909	2,612	49.4	49.1
Low (<= 12.6%)	60,492	49,528	6,353	52.0	51.8
Race (% White alone)					
High (> 73.3%)	53,713	43,455	4,828	52.7	52.6
Low (<= 73.3%)	25,827	22,982	4,136	48.8	48.5
Ethnicity (% Hispanic)					
High (> 17.3%)	21,063	18,589	2,944	49.4	49.3
Low (<= 17.3%)	58,477	47,848	6,020	52.1	51.9
Poverty status (% below poverty line)					
High (> 15.1%)	30,141	22,876	3,289	53.5	53.4
Low (<= 15.1%)	49,399	43,561	5,675	50.1	49.9
Census region					
Northeast	13,589	12,381	1,726	49.1	48.7
Midwest	18,866	14,242	1,494	54.5	54.4
South	29,458	25,365	3,635	50.4	50.2
West	17,627	14,449	2,109	51.6	51.5

[a] The unweighted response rate = $C_{HH}/(C_{HH}+N_{HH}+e_{HH}*U_{HH})$. The weighted response rate was calculated using a similar approach except that the Wave 4 household IPS weights were used to obtain weighted counts.

[b] The information used to define the subgroups was from the 5-year (2012-2016) ACS.

[22] Information from the 5-year (2012 to 2016) rather than the 1-year (2016) ACS was used because 1-year ACS estimates are not provided for smaller geographies such as census tracts or block groups.

As shown in Table B-1, the weighted overall response rate for the Wave 4 replenishment sample household screener was 51.2 percent. The weighted response rates of subgroups differed from one another by 5.0 percentage points for education, 4.1 percentage points for percent of the population identifying as White alone, 3.5 percentage points for poverty status, and 5.7 percentage points for census region. Smaller differences were observed for percent Black alone and percent Hispanic.

B.2 Shadow Youth Response Rates for Wave 4 Replenishment Sample

If a 10- or 11-year-old child was selected for the PATH Study at the conclusion of the household screener, the parent (or guardian) was asked to consent to their child joining the study as a shadow youth to be interviewed after turning 12. If the parent provided consent, as well as the child's date of birth, the child was considered a shadow youth respondent. If the parent did not provide consent, or did not provide the date of birth, the sampled child was treated as a shadow youth nonrespondent.

Table B-2 presents the weighted and unweighted shadow youth response rates for the Wave 4 replenishment sample. The shadow youth response rates (conditioning on completion of the household screener) were calculated as:

$$RR_{SY} = C_{SY}/(C_{SY} + N_{SY})$$

where

RR_{SY} = Wave 4 replenishment sample shadow youth response rate;

C_{SY} = number of Wave 4 shadow youth respondents; and

N_{SY} = number of Wave 4 shadow youth nonrespondents.

Unweighted response rates were calculated based on actual case counts, while weighted response rates were based on the sums of the Wave 4 IPS weights (calculated as the product of the household IPS weight and the inverse of the within-household probability of selection).

In addition to the overall row, the table includes rows on age, sex, and race/ethnicity based on information collected through the household screener, as well as census region. Shadow youth with missing values for a particular characteristic were excluded from the response rate calculation for that characteristic.

The decision to exclude shadow youth with missing values for a particular characteristic from the corresponding response rate calculation leads to results in Table B-2 that may seem inconsistent. For example, the weighted shadow youth response rate is 85.2 percent for 10- to 11-year-olds, and yet only 78.7 percent overall. The lower overall response rate reflects the study procedures applied to household members for whom no age information was provided by the household screener respondent. Such individuals were treated as shadow youth for sampling purposes; those selected were immediately designated as nonrespondents. These 305 cases appear in the row labeled "Age not reported" in Table B-2. Response rates (not shown) were also lower for shadow youth for whom sex, race, and/or ethnicity were not provided by the household screener respondent.

Table B-2. Wave 4 shadow youth response rates conditioning on household screener response (Wave 4 replenishment sample)

Characteristic reported by household screener respondent[a] or census region	C_{SY}: Shadow youth response (n)	N_{SY}: Shadow youth nonresponse (n)	RR_{SY}: Unweighted response rate[b] (%)	RR_{SY}: Weighted response rate[b] (%)
Overall	4,294	1,049	80.4	78.7
Age group				
10-11	4,273	744	85.2	85.2
Age not reported	0	305	0.0	0.0
Sex				
Male	2,249	442	83.6	82.9
Female	2,045	442	82.2	82.0
Race/ethnicity				
Hispanic	1,230	187	86.8	87.1
Non-Hispanic White alone	2,131	429	83.2	82.4
Non-Hispanic Black alone	479	111	81.2	81.2
Non-Hispanic other race or multiple races	440	112	79.7	79.7
Census region				
Northeast	619	176	77.9	76.2
Midwest	1,016	228	81.7	79.9
South	1,553	398	79.6	78.3
West	1,106	247	81.7	79.7

[a] The sum of the counts in all the categories may not be equal to the count in the overall row due to missing values or unusual circumstances. For example, the sum of the counts across age categories is not equal to the count in the overall row because for a small proportion of cases the age reported by the household screener respondent was outside of the 10-11 range.

[b] The unweighted response rate = $RR_{SY}/(C_{SY}+N_{SY})$. The weighted response rate was calculated using a similar approach except that the Wave 4 IPS weights were used to obtain weighted counts.

The weighted overall shadow youth response rate for the Wave 4 replenishment sample was 78.7 percent. The biggest subgroup difference occurred for race/ethnicity, with weighted response rates of 87.1 percent for Hispanic shadow youth and 79.7 percent for non-Hispanic shadow youth of

other or multiple races. Differences among subgroups on weighted response rates were 0.9 percentage points for sex and 3.7 percentage points for census region.