ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

ICPSR 37786

Population Assessment of Tobacco and Health (PATH) Study [United States] Special Collection 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

Wave 4.5 Data Collection Nonresponse Bias Analysis Report

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 4.5 Data Collection

Nonresponse Bias Analysis Report

March 2, 2020

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

Table of Contents

Chapter **Page**

Tables

Table of Contents

Tables (continued)

Figure

1. Introduction

Westat prepared this report as a reference document for researchers using the Population Assessment of Tobacco and Health (PATH) Study data. The PATH Study is a nationally representative, longitudinal cohort study of adults and youth in the United States, collecting data on those ages 12 years and older in each of the first four waves, and on those ages 12 to 17 in Wave 4.5. This report focuses on the Wave 4.5 response rates and potential nonresponse biases for select demographic and outcome measures. Similar reports are also available for Wave 1 through Wave 4, and can be found at the website for the National Addiction & HIV Data Archive Program (NAHDAP) (https://www.icpsr.umich.edu/icpsrweb/NAHDAP/studies/36231/datadocumentation).

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 (younger than 12 years old), 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 and Wave 4.5, the PATH Study data support analyses of two cohorts. 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 4.5. All Wave 1 respondents were eligible for Wave 4.5 data collection unless they were incarcerated, deceased, resided outside of the U.S., or were not 12 to 17 years old at the time of Wave 4.5. However, all interview respondents from the Wave 1 Cohort were at least 13 years of age at Wave 4.5.

The 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 both longitudinal and cross-sectional analyses at Wave 4.5.

This report summarizes the Wave 4.5 nonresponse bias analyses conducted separately for the two cohorts. The potential for nonresponse bias in Wave 4.5 estimates based on the Wave 1 Cohort is evaluated conditioning on Wave 1 response, while potential nonresponse bias in estimates based on the Wave 4 Cohort is assessed conditioning on Wave 4 response. With this in mind, a brief summary of the findings from the Wave 1 nonresponse bias analyses and from the Wave 4 nonresponse bias analyses for the Wave 4 Cohort are provided below.

Wave 1 Cohort Nonresponse Bias Analyses at Wave 1

- 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 weight, also known as the inverse-of-probability-of-selection (IPS) weight for the PATH Study. Exceptions were found for single-person households, sex, education, and ethnicity when comparing Wave 1 estimates using the IPS weight 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 cross-sectional weight, 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 weight. 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 weight. In addition, when the estimates were adjusted for interview nonresponse, they were found to approximate the ACS estimates more closely.

Wave 4 Cohort Nonresponse Bias Analyses at Wave 4

- When using the Wave 4 cross-sectional weight, the weighted demographic and socio-economic estimates for the Wave 4 adult interview for the Wave 4 Cohort were nearly identical to corresponding estimates based on the 2016 ACS. However, adults with health insurance were slightly underrepresented by the Wave 4 Cohort weighted estimate. Weighted estimates of cigarette-smoking rates among adults in the Wave 4 Cohort at Wave 4 were within the range of estimates found in other national health studies.

- Weighted demographic estimates based on all Wave 4 Cohort youth at Wave 4 were nearly identical to corresponding estimates based on the 2016 ACS. The PATH Study ever-tried cigarette-smoking rates for all Wave 4 Cohort youth in Wave 4 remained 1 to 6 percentage points lower than those found in other national studies, when using the Wave 4 cross sectional weight.

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 Wave 4.5 youth interview 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 were conducted with respondents ages 12 years and older in each of the first four waves and with those ages 12 to 17 in Wave 4.5. 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, in the military, or in 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

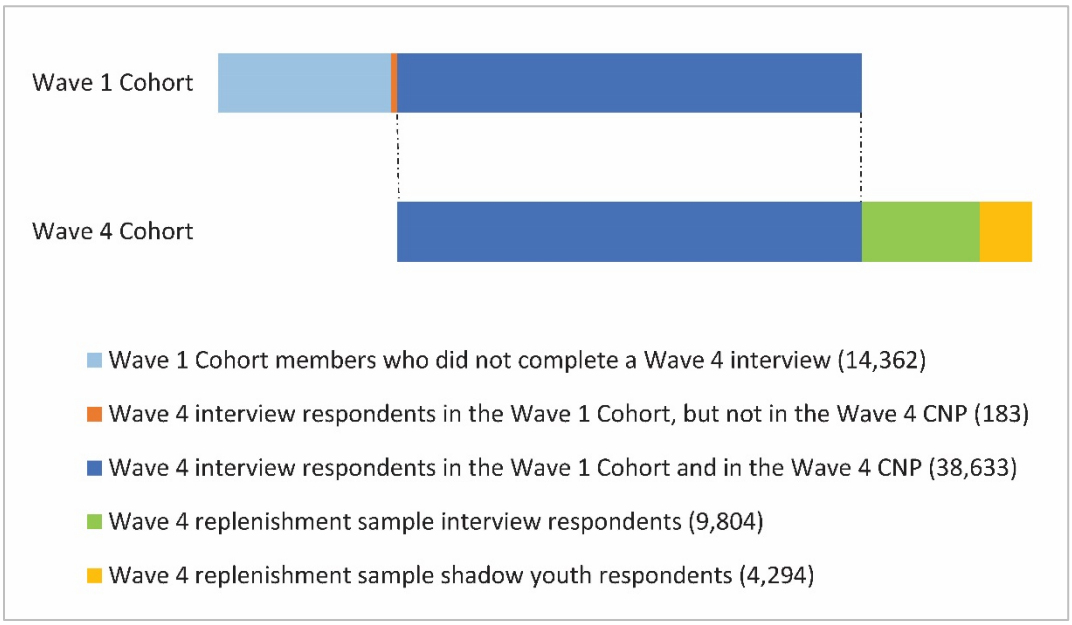

Table 2-1. Wave 1 Cohort and Wave 4 Cohort sample sizes

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 (RUFs) 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.

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

[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 (e.g., twins), 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.

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 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 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 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

[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.

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

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.

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 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. Among the Wave 1 Cohort members, 27,757 adult interviews and 11,059 youth interviews were completed at Wave 4. Among Wave 4 Cohort members (i.e., after combining the Wave 1 Cohort and Wave 4 replenishment samples and restricting to those 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 Wave 4 Cohort

interview respondents, the study collected 20,928 urine and 3,602 blood specimens from adults and 13,095 urine specimens from youth. The Wave 4 Data and Biospecimen Collection Report is available on the NAHDAP website.

2.4 Wave 4.5

Wave 4.5 was the first special data collection of the PATH Study; its focus was on continuing the annual collection of data among youth ages 12 to 17 (and their parents). There was no additional sampling for Wave 4.5 and no biospecimens were collected.

Wave 4.5 of the PATH Study was the fourth follow-up wave for age-eligible members of the Wave 1 Cohort and the first follow-up wave for age-eligible participants selected as part of the Wave 4 replenishment sample. All study participants were eligible for Wave 4.5 if they were ages 12 to 17, still residents of the U.S., and not incarcerated. Nonrespondents from previous waves were asked to participate in Wave 4.5 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 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 any two consecutive waves were not fielded for data collection.[6] Shadow youth from previous waves who turned age 12 by Wave 4.5 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 4.5.

The PATH Study Wave 4.5 interview data collection began December 1, 2017 and ended December 1, 2018. The PATH Study completed 13,131 youth interviews at Wave 4.5; 8,761 of these youth were members of the Wave 1 Cohort and 12,918 were members of the Wave 4 Cohort.

A summary of the results of the data and biospecimen collection efforts for Waves 1 to 4.5, by cohort and age group, is provided in Table 2-2.

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

Table 2-2. PATH Study data collection summary by wave, cohort, and age group

Wave	Data collection start date	Data collection end date	Cohort	Age group	Interviews conducted	Urine specimens collected[a]	Blood specimens collected[b]
1	September 12, 2013	December 14, 2014	Wave 1 Cohort	Adults	32,320	21,801	14,520
				Youth	13,651	N/A	N/A
2	October 23, 2014	October 30, 2015	Wave 1 Cohort	Adults	28,362	13,696	908
				Youth	12,172	N/A	N/A
3	October 19, 2015	October 23, 2016	Wave 1 Cohort	Adults	28,148	14,979	835
				Youth	11,814	N/A	N/A
4	December 1, 2016	January 3, 2018	Wave 1 Cohort	Adults	27,757	16,131	890
				Youth	11,059	9,892	N/A
			Wave 4 Cohort	Adults	33,644	20,928	3,602
				Youth	14,793	13,095	N/A
4.5	December 1, 2017	December 1, 2018	Wave 1 Cohort	Youth	8,761	N/A	N/A
			Wave 4 Cohort	Youth	12,918	N/A	N/A

[a] The PATH Study only requested urine specimens from first-time adult interview respondents, a subsample of adults who had completed an adult interview at a prior wave, and all youth at Wave 4.

[b] The PATH Study only requested blood specimens from first-time adult interview respondents.

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 4.5 Potential Nonresponse Bias

It is not always possible to measure the actual bias due to nonresponse; however, various approaches can be used to help identify potential sources of nonresponse bias. At Wave 4.5, the Wave 1 Cohort data serve the purpose of longitudinal analysis, while the Wave 4 Cohort data support both cross-sectional and longitudinal analyses. Separate sets of analyses were conducted for the two cohorts to assess potential nonresponse bias in estimates using data from Wave 4.5 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 4.5 response rates were calculated for the Wave 1 Cohort and the Wave 4 Cohort 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).

For the Wave 1 Cohort, the Wave 4.5 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 for Waves 2 through 4. In contrast, the Wave 4.5 youth interview response rates for the Wave 4 Cohort are conditioning on Wave 4 response, so the subgroups are based on Wave 4 information, apart from the recruitment wave subgroup (which distinguishes Wave 4 Cohort members who were recruited at Wave 1 from those recruited at Wave 4). This should be kept in mind when reviewing the response rates in Sections 4.1 and 5.1.

A Wave 4.5 nonrespondent does not have a Wave 4.5 interview date, so his/her Wave 4.5 age was determined using the latest available date of birth or age information. Each study participant fielded

for Wave 4.5 was assigned an "anniversary month" for Wave 4.5, as described in Section 3.5 of the PATH Study RUFs User Guide. The age classification date for a Wave 4.5 nonrespondent was 1 month after the last day of his/her anniversary month or the final date of the Wave 4.5 data collection, whichever was earlier. A nonrespondent was considered eligible for a Wave 4.5 youth interview if his/her age was determined to be between 12 and 17 on the age classification date.

Both weighted and unweighted response rates are reported for the youth interview. 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. The basic design weight associated with the Wave 1 sample selection, also known as the Wave 1 IPS weight, was used for calculating the Wave 1 Cohort weighted response rates. A Wave 4 IPS weight does not exist for the Wave 4 Cohort. 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 full 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, the Wave 4 cross-sectional weight was used for calculating the Wave 4 Cohort weighted response rates.

3.2.2 Comparison to "Frame" Data

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 and Wave 4 Cohorts, separately, estimates based on Wave 4.5 youth interview respondents were

compared to the "frame" information from the cohort's baseline wave (i.e., Wave 1 or Wave 4) for Wave 4.5 youth. For the Wave 1 Cohort, this comparison helps identify characteristics that might be associated with nonresponse bias due to attrition between Wave 1 and Wave 4.5, after compensating for possible undercoverage of the target population at Wave 1 (due to sampling frame deficiencies) and Wave 1 nonresponse. For the Wave 4 Cohort, the comparison helps identify characteristics that might be associated with nonresponse bias due to attrition between Wave 4 and Wave 4.5, after compensating for d possible undercoverage and Wave 4 nonresponse.

Wave 1 Cohort

For the Wave 1 Cohort, the "frame" information included Wave 1 demographic 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 4.5 interview and were computed using the Wave 1 cross-sectional weight (which was designed to reduce potential nonresponse bias at Wave 1).

Two sets of Wave 1 Cohort estimates were obtained from Wave 4.5 interview respondents. The first set of estimates was based on the Wave 1 cross-sectional weight. Comparing this set of estimates to the "frame" data shows the representativeness of the Wave 4.5 responding sample before any statistical adjustment for attrition between Waves 1 and 4.5. The second set of estimates was based on the Wave 4.5 single-wave weight that applies to Wave 4.5 respondents in the Wave 1 Cohort regardless of their Wave 2, Wave 3, or Wave 4 response status (referred to as the "single-wave weight"). The Wave 4.5 single-wave weight for the Wave 1 Cohort was designed to reduce potential nonresponse bias due to attrition between Waves 1 and 4.5, so the second set of estimates shows the extent to which the statistical weighting adjustment might have helped reduce potential nonresponse bias at Wave 4.5. For reference, the terms "before Wave 4.5 weighting adjustment" and "after Wave 4.5 weighting adjustment" are used to refer to the two sets of estimates and comparisons in Section 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 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 RUFs User Guide.

The final person-level weights assigned to Wave 1 respondents, i.e., the Wave 1 cross-sectional weights, served as the initial weights for developing the Wave 1 Cohort single-wave weight for Wave 4.5 interview respondents. These weights were adjusted to account for nonresponse to the Wave 4.5 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 4.5 single-wave weight construction for the Wave 1 Cohort appear in Section 5.1.4 of the PATH Study RUFs User Guide.

Wave 4 Cohort

For the Wave 4 Cohort, the "frame" information included Wave 4 demographic characteristics, as well as Wave 4 tobacco-use status. The estimates were based on all the Wave 4 Cohort members who were eligible for the Wave 4.5 interview and were computed using the Wave 4 cross-sectional weight (which was designed to reduce potential nonresponse bias at Wave 4).

Two sets of Wave 4 Cohort estimates were obtained from Wave 4.5 interview respondents: "before Wave 4.5 weighting adjustment" and "after Wave 4.5 weighting adjustment." The first set of estimates was based on the Wave 4 cross-sectional weight. Comparing this set of estimates to the "frame" data shows the representativeness of the Wave 4.5 responding sample before any statistical adjustment for attrition between Waves 4 and 4.5. The second set of estimates was based on the Wave 4.5 single-wave weight for the Wave 4 Cohort, and shows the extent to which the statistical weighting adjustment might have helped reduce potential nonresponse bias at Wave 4.5.

To create the Wave 4 cross-sectional weight for the Wave 4 Cohort, a sequence of weighting steps involving the Wave 1 Cohort single-wave weight and the Wave 4 replenishment sample weight was implemented, including nonresponse adjustments, preliminary raking, compositing, and final raking and trimming. The control totals were created using the 2016 ACS Public Use Microdata Sample (PUMS). For shadow youth and youth, raking was done using cross-classifications of census region, single-year of age, race/ethnicity, and sex. (Missing values in the variables used for raking were imputed.)

The approach for the construction of the Wave 4.5 single-wave weight for the Wave 4 Cohort was similar to that for the Wave 1 Cohort, except that the Wave 4 cross-sectional weights served as the initial weights and the control totals either came from the 2016 ACS or were sample-based and reflected estimated Wave 4 population characteristics. More details about the construction of the Wave 4 cross-sectional weight and the Wave 4.5 single-wave weight for the Wave 4 Cohort appear in Sections 5.1.4 and 5.1.5, respectively, of the PATH Study RUFs 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 can be used to produce cross-sectional estimates (which are approximate in nature for Waves 2 and 3). For Waves 4 and 4.5, the

Wave 4 Cohort serves the purpose of cross-sectional analysis (which is approximate in nature for Wave 4.5). 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).

Estimates were obtained using the Wave 4.5 youth interview data and Wave 4.5 single-wave weight for the Wave 4 Cohort. Several PATH Study measures of cigarette smoking and electronic nicotine device systems (ENDS)[7] use were compared to those from other national studies of tobacco and health (for the most similar timeframes for which data were available). The PATH Study collects data on a range of tobacco-use behaviors; many of these variables are not available in other studies. However, responses to the PATH Study questions on cigarette smoking and ENDS use can be compared with estimates from other national studies that ask about these behaviors.

The external data sources for cigarette smoking included the National Health and Nutrition Examination Survey 2015-2016 (NHANES 2015-2016), the National Survey on Drug Use and Health 2017 (NSDUH 2017), and the National Youth Tobacco Survey 2018 (NYTS 2018). The cigarette-smoking data for all three external studies were restricted to ages 12-17 to match the Wave 4.5 age eligibility criterion. The external data sources used for comparison of ENDS use included NYTS 2018[8] and the National Youth Risk Behavior Survey 2017 (NYRBS 2017). NYRBS is a survey of only high school students. Therefore, the PATH Study, NYTS, and NYRBS estimates of ENDS use were restricted to ages 14-17. Appendices A and B describe the questions used to

[7] While the term "electronic nicotine products" is used in the PATH Study instrument to represent all electronic nicotine devices such as e-cigarettes, e-cigars and e-hookahs, this report uses the more technical term ENDS because it is more commonly used in the tobacco-use literature.

[8] NYTS 2018 asks about e-cigarette use and not ENDS use. However, the definition of e-cigarettes provided to NYTS respondents just prior to asking about e-cigarette use frequency is virtually identical to the definition of "electronic nicotine products" provided to PATH Study respondents. Specifically, the PATH Study includes e-cigarettes as a specific kind of "electronic nicotine product," whereas the NYTS defines e-cigarettes as encompassing all ENDS use (e.g., e-cigarette, e-cigar, e-hookah, e-pipe). With these definitions in mind, NYTS 2018 e-cigarette use estimates are equivalent to ENDS use estimates from Wave 4.5 of the PATH Study and NYRBS 2017 in Table 5.5.

define cigarette use and ENDS use on these surveys as well as the PATH Study, and outlines differences in their target populations.

For both the PATH Study and the external data sources for cigarette smoking and ENDS use, 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 in both the PATH Study and the studies that were used for comparison purposes. Specifically, the item-missing data rates were generally below 3.1 percent for youth cigarette and ENDS estimates at the overall level, with the exception of 5.9 percent missing for youth ever tried cigarette smoking e from the NHANES 2015-2016 data and 13 percent missing for youth ENDS use in the last 30 days from the NYRBS 2017. [9]

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/STAT software version 15.1 was used to calculate all the point estimates. Confidence intervals were estimated using the modified Wilson approach (Wilson, 1927; SAS Institute, 2018). 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 and the second, more conservative approach was used to compare estimates between the PATH Study and external sources in Section 5.3 and 5.4. No adjustments were made for multiple comparisons

[9] However, note that the NSDUH 2017 public use file included only "usable" cases that met specified minimum item response requirements.

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 4.5

This chapter presents results of nonresponse bias analyses for the Wave 1 Cohort at Wave 4.5. The interview nonresponse bias analyses for the Wave 1 Cohort at Wave 4.5 are all conditioning on Wave 1 response. Section 4.1 describes how the youth interview response rate varies across select subgroups. Section 4.2 compares demographic characteristics as well as tobacco-use status between Wave 4.5 youth interview respondents and those eligible for the Wave 4.5 youth interview.

4.1 Youth Interview Response Rates Conditioning on Wave 1 Response

Among those in the Wave 1 Cohort, unweighted and weighted response rates were calculated for the Wave 4.5 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 4.5 youth interview response rate;

C_Y = number of Wave 4.5 youth interview complete cases;

N_Y = number of Wave 4.5 youth interview nonrespondents known to be eligible;

U_Y = number of Wave 4.5 youth interview nonrespondents with unknown eligibility status;

e_Y = estimated proportion of eligible cases among the Wave 4.5 youth interview nonrespondents with unknown eligibility status; and

I_Y = number of Wave 4.5 youth interview ineligible cases that were not permanently ineligible prior to Wave 4.5.

Unweighted counts and weighted counts based on the Wave 1 IPS weight 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 4.5

due to permanent incarceration, death, or having permanently moved overseas; this set of cases, denoted as IP_Y in Table 4-1, were not included in the equation for computing e_Y.

Table 4-1 provides response rates for the Wave 4.5 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, census region, 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 4.5 youth interview were 74.5 and 74.6 percent, respectively; the unweighted and weighted response rates for each subgroup were also very similar to each other. The response rates varied little by sex (e.g., male versus female), but to a greater extent by Wave 1 age, race/ethnicity, and census region, and substantially by earlier waves' response status. The Wave 4.5 youth interview weighted response rate for those who responded in Waves 2, 3, and 4 was 93.3 percent; for those who skipped at least one of these waves the weighted response rate was 19.3 percent. For Wave 1 age, the weighted response rate for the "under 12" age group (72.4 percent) was noticeably lower than for the "12-13" age group (77.2 percent). The weighted response rate for the "14-17" age group (92.4 percent) is based on a relatively small sample size.[10] Response rates also differed somewhat across race/ethnicity groups, with the highest among non-Hispanic Black alone youth (77.4 percent) and lowest among non-Hispanic White alone youth (72.6 percent). By Wave 1 census region, the West had the highest response rate (76.8 percent) and the Northeast had the lowest (71.5 percent). For tobacco use, the weighted response rate among Wave 1 youth was 4.2 percentage points lower for the Wave 1 "ever user" group than for the Wave 1 "never user" group.[11]

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 4.5 youth interview estimates.

[10] In theory, all participants who were classified in the "14-17" age group at Wave 1 should have reached adulthood by Wave 4.5. The 215 participants counted in the "14-17" age group in Table 4-1 are the result of age corrections since Wave 1.

[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 were nonsmokers.

Table 4-1. Wave 4.5 youth interview response rates conditioning on Wave 1 response

Characteristic[a]	C_Y: Completed (n)	IP_Y: Permanently ineligible prior to Wave 4.5 (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	8,761	20	91	1,353	1,665	74.5	74.6
Earlier waves' response status							
Responded in all earlier waves	8,202	0	54	505	86	93.3	93.3
Did not respond in at least one earlier wave	559	20	37	848	1,579	19.0	19.3
Wave 1 sex							
Male	4,540	9	50	694	860	74.6	74.8
Female	4,206	10	40	658	799	74.4	74.5
Wave 1 age group							
Under 12	5,180	13	18	912	1,082	72.2	72.4
12-13	3,395	7	58	430	579	77.2	77.2
14-17	186	0	15	11	3	93.1	92.4
Wave 1 race/ethnicity[b]							
Hispanic	2,584	12	29	337	444	76.9	77.0
Non-Hispanic White alone	4,058	1	27	679	872	72.4	72.6
Non-Hispanic Black alone	1,222	1	18	186	174	77.4	77.4
Non-Hispanic other race or multiple races	785	6	13	131	154	73.5	73.7
Wave 1 census region							
Northeast	1,201	5	12	200	283	71.4	71.5
Midwest	1,960	4	13	287	394	74.3	74.6
South	3,268	1	39	546	599	74.2	74.2
West	2,332	10	27	320	389	76.8	76.8
Wave 1 ever tobacco use[c] (Wave 1 youth only)							
Ever user	259	1	11	47	43	74.5	74.5
Never user	3,127	6	57	365	491	78.7	78.7

[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] Hispanic refers to Hispanic or Latino; Black refers to Black or African-American.

[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.

4.2 Comparison Between Wave 4.5 Respondents and Those Eligible for the Wave 4.5 Interview (Among the Wave 1 Cohort)

This section compares weighted estimates between the Wave 4.5 youth interview respondents in the Wave 1 Cohort and the Wave 1 respondents who were eligible for the Wave 4.5 youth interview. The weighted estimates cover demographic characteristics (shown in Table 4-2) and tobacco-use status (shown in Table 4-3).

Section 3.2.2 describes the methods used for obtaining the results in Tables 4-2 and 4-3. Each table shows two sets of comparisons between the Wave 4.5 youth interview respondents and those eligible for the Wave 4.5 youth interview. The first set is for the "before Wave 4.5 weighting adjustment" comparison, and the second set is for the "after Wave 4.5 weighting adjustment" comparison.

Table 4-2 compares Wave 1 demographic characteristics between the Wave 4.5 youth interview respondents in the Wave 1 Cohort and those Wave 1 respondents eligible for the Wave 4.5 youth interview. In Table 4-2, the percentages sum to 100 percent[12] over the categories associated with each characteristic. For example, for both the Wave 4.5 respondents and those eligible for the Wave 4.5 youth interview, the estimated percentage of males and the estimated percentage of females add to 100 percent.

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. Table 4-3 has two columns labeled "Sample size." These columns represent the denominator for calculating an unweighted estimate of ever tobacco use for the category associated with the characteristic. These definitions generalize to all tables in this report that use a column heading of "Unweighted count" or "Sample size."

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

Table 4-2. Comparison of Wave 1 demographic characteristics between Wave 1 respondents who were eligible for Wave 4.5 youth interview and Wave 4.5 youth interview respondents

Wave 1 characteristic	Wave 1 respondents who were eligible for Wave 4.5 youth interview		Wave 1 respondents who completed Wave 4.5 youth interview				
				Before Wave 4.5 weighting adjustment		After Wave 4.5 weighting adjustment	
	Unweighted count	A: Weighted percentage, using Wave 1 cross-sectional weight [95% confidence interval]	Unweighted count	B: Weighted percentage, using Wave 1 cross-sectional weight [95% confidence interval]	Difference in weighted percentages [B – A] [95% confidence interval]	C: Weighted percentage, using Wave 4.5 single-wave weight [95% confidence interval]	Difference in weighted percentages [C – A] [95% confidence interval]
Sex							
Male	6,094	51.3% [50.4%, 52.2%]	4,540	51.4% [50.4%, 52.5%]	0.1% [-0.5%, 0.8%]	51.3% [50.3%, 52.4%]	0.0% [-0.1%, 0.1%]
Female	5,663	48.7% [47.8%, 49.6%]	4,206	48.6% [47.5%, 49.6%]	-0.1% [-0.8%, 0.5%]	48.7% [47.6%, 49.7%]	0.0% [-0.1%, 0.1%]
Age group							
Under 12	7,174	59.7% [58.8%, 60.6%]	5,180	58.0% [56.9%, 59.0%]	-1.8% [-2.3%, -1.2%]	59.8% [58.7%, 60.8%]	0.0% [0.0%, 0.1%]
12-13	4,404	38.5% [37.6%, 39.4%]	3,395	39.8% [38.8%, 40.9%]	1.3% [0.8%, 1.9%]	38.4% [37.4%, 39.5%]	0.0% [-0.1%, -0.0%]
14-17	200	1.8% [1.6%, 2.0%]	186	2.2% [1.9, 2.5%]	0.4% [0.3%, 0.5%]	1.8% [1.5%, 2.1%]	0.0% [-0.0%, -0.0%]
Race/ethnicity[a]							
Hispanic	3,365	23.4% [22.7%, 24.2%]	2,584	24.1% [23.2%, 25.0%]	0.7% [0.1%, 1.3%]	23.4% [22.5%, 24.3%]	-0.0% [-0.1%, 0.1%]
Non-Hispanic White alone	5,609	53.2% [52.3%, 54.1%]	4,058	52.1% [51.0%, 53.1%]	-1.2% [-1.9%, -0.5%]	53.2% [52.1%, 54.2%]	-0.1% [-0.2%, 0.0%]
Non-Hispanic Black alone	1,582	13.6% [13.0%, 14.2%]	1,222	14.0% [13.3%, 14.8%]	0.4% [0.0%, 0.9%]	13.6% [12.9%, 14.4%]	0.1% [-0.0%, 0.1%]
Non-Hispanic other race or multiple races	1,070	9.7% [9.2%, 10.3%]	785	9.8% [9.2%, 10.5%]	0.1% [-0.3%, 0.5%]	9.8% [9.2%, 10.4%]	0.0% [-0.0%, 0.1%]

Table 4-2. Comparison of Wave 1 demographic characteristics between Wave 1 respondents who were eligible for Wave 4.5 youth interview and Wave 4.5 youth interview respondents (continued)

Wave 1 characteristic	Wave 1 respondents who were eligible for Wave 4.5 youth interview		Wave 1 respondents who completed Wave 4.5 youth interview				
				Before Wave 4.5 weighting adjustment		After Wave 4.5 weighting adjustment	
	Unweighted count	A: Weighted percentage, using Wave 1 cross-sectional weight [95% confidence interval]	Unweighted count	B: Weighted percentage, using Wave 1 cross-sectional weight [95% confidence interval]	Difference in weighted percentages [B – A] [95% confidence interval]	C: Weighted percentage, using Wave 4.5 single-wave weight [95% confidence interval]	Difference in weighted percentages [C – A] [95% confidence interval]
Census region							
Northeast	1,684	16.3% [15.7%, 17.0%]	1,201	15.8% [15.0%, 16.6%]	-0.6% [-1.0%, -0.1%]	16.3% [15.6%, 17.1%]	0.0% [-0.1%, 0.0%]
Midwest	2,641	22.0% [21.2%, 22.7%]	1,960	22.1% [21.3%, 23.0%]	0.2% [-0.4%, 0.7%]	22.0% [21.1%, 22.9%]	0.0% [-0.0%, 0.1%]
South	4,413	37.8% [37.0%, 38.7%]	3,268	37.4% [36.4%, 38.4%]	-0.4% [-1.2%, 0.4%]	37.8% [36.8%, 38.8%]	0.0% [-0.1%, 0.0%]
West	3,041	23.9% [23.1%, 24.7%]	2,332	24.7% [23.8%, 25.6%]	0.8% [0.2%, 1.4%]	23.9% [23.0%, 24.8%]	0.0% [-0.0%, 0.1%]

[a] Hispanic refers to Hispanic or Latino; Black refers to Black or African-American.

Table 4-3. Comparison of Wave 1 "ever tobacco-use" estimates between Wave 1 youth who were eligible for Wave 4.5 youth interview and Wave 1 youth who completed Wave 4.5 youth interview*

Wave 1 characteristic[a]	Wave 1 youth who were eligible for Wave 4.5 youth interview		Wave 4.5 youth interview respondents who completed Wave 1 youth interview				
				Before Wave 4.5 weighting adjustment		After Wave 4.5 weighting adjustment	
	Sample size	A: Weighted estimate, using Wave 1 cross-sectional weight [95% confidence interval]	Sample size	B: Weighted estimate, using Wave 1 cross-sectional weight [95% confidence interval]	Difference in weighted estimates [B – A] [95% confidence interval]	C: Weighted estimate, using Wave 4.5 single-wave weight [95% confidence interval]	Difference in weighted estimates [C – A] [95% confidence interval]
Overall	4,332	7.6% [6.6%, 8.7%]	3,386	7.1% [6.0%, 8.4%]	-0.5% [-0.9%, -0.0%]	7.2% [6.1%, 8.5%]	-0.4% [-0.8%, -0.1%]
Sex							
Male	2,207	8.7% [7.3%, 10.2%]	1,725	8.2% [6.7%, 10.0%]	-0.5% [-1.1%, 0.1%]	8.3% [6.8%, 10.1%]	-0.4% [-1.0%, 0.3%]
Female	2,108	6.4% [5.2%, 7.9%]	1,648	6.0% [4.6%, 7.6%]	-0.5% [-1.1%, 0.2%]	6.0% [4.7%, 7.6%]	-0.4% [-1.0%, 0.2%]
Age group[b]							
12-13	4,142	7.5% [6.5%, 8.6%]	3,210	7.0% [5.9%, 8.3%]	-0.5% [-0.9%, 0.0%]	7.1% [6.0%, 8.4%]	-0.3% [-0.8%, 0.1%]
14-17	190	9.5% [6.0%, 14.9%]	176	8.6% [5.3%, 13.7%]	-0.9% [-2.8%, 0.9%]	8.6% [5.3%, 13.7%]	-0.9% [-2.7%, 0.9%]
Race/ethnicity[c]							
Hispanic	1,257	7.3% [5.9%, 9.0%]	1,028	6.7% [5.3%, 8.5%]	-0.6% [-1.4%, 0.2%]	6.6% [5.2%, 8.3%]	-0.7% [-1.6%, 0.1%]
Non-Hispanic White alone	1,991	8.3% [6.9%, 10.0%]	1,504	7.5% [5.9%, 9.6%]	-0.8% [-1.6%, -0.0%]	7.6% [6.0%, 9.7%]	-0.7% [-1.5%, 0.0%]
Non-Hispanic Black alone	566	5.2% [3.2%, 8.3%]	456	6.0% [3.7%, 9.6%]	0.8% [0.1%, 1.6%]	6.1% [3.8%, 9.8%]	1.0% [0.1%, 1.9%]
Non-Hispanic other race or multiple races	393	7.9% [5.7%, 11.0%]	302	8.3% [5.7%, 12.0%]	0.4% [-0.9%, 1.7%]	8.8% [6.1%, 12.6%]	0.9% [-0.4%, 2.3%]

Table 4-3. Comparison of Wave 1 "ever tobacco-use" estimates between Wave 1 youth who were eligible for Wave 4.5 youth interview and Wave 1 youth who completed Wave 4.5 youth interview * (continued)

Wave 1 characteristic[a]	Wave 1 youth who were eligible for Wave 4.5 youth interview		Wave 4.5 youth interview respondents who completed Wave 1 youth interview				
				Before Wave 4.5 weighting adjustment		After Wave 4.5 weighting adjustment	
	Sample size	A: Weighted estimate, using Wave 1 cross-sectional weight [95% confidence interval]	Sample size	B: Weighted estimate, using Wave 1 cross-sectional weight [95% confidence interval]	Difference in weighted estimates [B – A] [95% confidence interval]	C: Weighted estimate, using Wave 4.5 single-wave weight [95% confidence interval]	Difference in weighted estimates [C – A] [95% confidence interval]
Census region							
Northeast	608	4.7% [3.2%, 6.9%]	449	4.6% [2.8%, 7.5%]	-0.1% [-1.1%, 0.8%]	4.6% [2.8%, 7.6%]	-0.1% [-1.2%, 0.9%]
Midwest	969	8.5% [6.2%, 11.6%]	760	8.6% [6.0%, 12.2%]	0.1% [-0.8%, 1.1%]	8.8% [6.1%, 12.4%]	0.3% [-0.7%, 1.3%]
South	1,632	8.2% [6.5%, 10.2%]	1,254	7.3% [5.4%, 9.9%]	-0.8% [-1.8%, 0.1%]	7.4% [5.5%, 10.0%]	-0.8% [-1.7%, 0.2%]
West	1,123	7.7% [6.2%, 9.7%]	923	7.0% [5.3%, 9.0%]	-0.8% [-1.4%, -0.1%]	7.2% [5.5%, 9.4%]	-0.6% [-1.3%, 0.1%]

* 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 4-3 includes a subset of the youth in Table 4-2 because there are no Wave 1 interview data for the Wave 4.5 youth interview respondents who were shadow youth at Wave 1.

[c] Hispanic refers to Hispanic or Latino; Black refers to Black or African-American.

For sex, the "before Wave 4.5 weighting adjustment" distribution among the Wave 4.5 youth interview respondents was almost the same as the distribution among those eligible for the Wave 4.5 youth interview. Non-Hispanic White alone youth as well as youth from the Northeast census region were slightly underrepresented among the Wave 4.5 respondents, while Hispanic youth, non-Hispanic Black alone youth, and youth from the West were slightly overrepresented. The most noticeable underrepresentation among the Wave 4.5 youth respondents was the Wave 1 "under 12" age group (with corresponding slight overrepresentation between the two older age groups).

For the "after Wave 4.5 weighting adjustment" differences, the estimated confidence intervals (in the last column of Table 4-2) either included zero or were in the proximity of zero for each of the Wave 1 characteristics examined: sex, age, race/ethnicity, and census region.

Table 4-3 shows the comparison of Wave 1 "ever tobacco-use" rates between Wave 1 youth interview respondents who completed the Wave 4.5 youth interview and all the Wave 1 youth interview respondents eligible for the Wave 4.5 youth interview. (Table 4-3 includes a subset of the youth in Table 4-2 because there are no Wave 1 interview data for the Wave 4.5 youth interview respondents who were shadow youth at Wave 1.) Besides the overall estimates, these tables also show the estimated Wave 1 "ever tobacco-use" rates for some demographic subgroups.

For both the "before Wave 4.5 weighting adjustment" and the "after Wave 4.5 weighting adjustment" estimates, differences between respondents and those eligible for the interview were no more than 1.0 percent. The largest difference was for the Wave 1 "ever tobacco-use" measure among non-Hispanic Black alone youth. However, most of the confidence intervals either included or were in the proximity of zero.

Overall, these results reflect the use of both demographic variables and tobacco-use measures from Wave 1 for calibrating the Wave 4.5 single-wave weight for the Wave 1 Cohort. Assuming that the Wave 1 demographic and tobacco-use characteristics in Tables 4-2 and 4-3 are correlated with key tobacco- and health-related outcome measures in Wave 4.5 of the PATH Study, 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 4.5.

5. Results of Wave 4 Cohort Nonresponse Bias Analyses at Wave 4.5

This chapter presents results of nonresponse bias analyses for the Wave 4 Cohort. 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 4.5 was the fourth follow-up attempt for those sampled at Wave 1, whereas members of the Wave 4 replenishment sample were asked to participate in the PATH Study for the second time. Section 5.1 describes how the youth interview response rate varies across select subgroups, including recruitment wave. Section 5.2 compares demographic characteristics as well as tobacco-use status between Wave 4.5 youth interview respondents and those eligible for the Wave 4.5 youth interview. Sections 5.3 and 5.4 compare cross-sectional estimates of cigarette smoking and ENDS use between the PATH Study Wave 4 Cohort at Wave 4.5 and external sources.

5.1 Youth Interview Response Rates Conditioning on Wave 4 Response (Among the Wave 4 Cohort)

This section summarizes the Wave 4 Cohort response rates for the Wave 4.5 youth interview, conditioning on Wave 4 response. Similar to what was described in Section 4.1 for the Wave 1 Cohort, unweighted and weighted response rates were calculated for the Wave 4.5 youth interview among those in the Wave 4 Cohort 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 4.5 youth interview response rate;

C_Y = number of Wave 4.5 youth interview complete cases;

N_Y = number of Wave 4.5 youth interview nonrespondents known to be eligible;

U_Y = number of Wave 4.5 youth interview nonrespondents with unknown eligibility status;

e_Y = estimated proportion of eligible cases among the Wave 4.5 youth interview nonrespondents with unknown eligibility status; and

I_Y = number of Wave 4.5 youth interview ineligible cases.[13]

Unweighted counts and weighted counts based on the Wave 4 cross-sectional weight were obtained for response status categories C_Y, N_Y, U_Y, and I_Y for unweighted response rates and weighted response rates, respectively. Ideally, weighted response rates for the Wave 4 Cohort would use an IPS weight, as has been the practice for Wave 1 Cohort weighted response rates. However, the information needed to create the IPS weight for the Wave 4 Cohort is not available (see Section 3.2.1). The Wave 4 cross-sectional weight was used for the weighted response rates presented in Table 5-1 because the sum of these weights provides the closest approximation to the Wave 4 Cohort target population at Wave 4, i.e., the U.S. CNP ages 10 and above as of Wave 4.

Table 5-1 provides response rates for the Wave 4.5 youth interview. In addition to the overall row, response rates are shown by Wave 4 sex, age, race/ethnicity, census region, tobacco-use status, and recruitment wave. 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 4.5 youth interview were 89.6 percent and 89.1 percent, respectively; the unweighted and weighted response rates for each subgroup were also similar to each other. The response rates varied modestly by Wave 4 sex, race/ethnicity, and census region. However, for Wave 4 age, the weighted response rate for the "under 12" age group (78.9 percent) was noticeably lower than for the "12-13" and "14-17" age groups (91.9 percent and 90.5 percent, respectively). This is likely because the persons in the "under 12" age group were shadow youth at Wave 4, so the Wave 4.5 youth interview was the first PATH Study interview they were asked to complete. (A similar pattern was observed among the Wave 1 Cohort.) For tobacco use, the weighted response rate among Wave 4 Cohort youth was 1.5 percentage points lower for the Wave 4 "ever user" group than for the Wave 4 "never user" group.

[13] Only youth fielded for Wave 4.5 and found to be ineligible are included in this set of cases for response rate calculations. Study participants not fielded on the basis of their age at prior waves are excluded.

Table 5-1. Wave 4.5 youth interview response rates conditioning on Wave 4 response (Wave 4 Cohort)

Characteristic[a]	C_Y: Completed (n)	I_Y: 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	12,918	107	1,252	246	89.6	89.1
Wave 4 sex						
Male	6,707	60	645	127	89.7	89.2
Female	6,170	46	603	119	89.5	88.9
Wave 4 age group						
Under 12	1,660	35	344	69	80.1	78.9
12-13	4,373	6	307	74	92.0	91.9
14-17	6,885	66	601	103	90.7	90.5
Wave 4 race/ethnicity[b]						
Hispanic	3,769	33	308	92	90.4	89.5
Non-Hispanic White alone	5,831	32	603	82	89.5	89.2
Non-Hispanic Black alone	1,615	20	159	39	89.1	87.8
Non-Hispanic other race or multiple races	1,214	16	138	25	88.2	88.6
Wave 4 census region						
Northeast	1,814	15	203	21	89.0	88.8
Midwest	2,868	14	285	50	89.5	89.3
South	4,841	50	476	115	89.1	88.5
West	3,395	28	288	60	90.7	89.9
Wave 4 ever tobacco use[c] (Wave 4 youth only)						
Ever user	1,853	29	176	34	89.8	90.0
Never user	8,841	38	672	126	91.7	91.5
Recruitment wave						
Wave 1	8,548	55	554	104	92.9	92.6
Wave 4	4,370	52	698	142	83.9	83.2

[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.

[b] Hispanic refers to Hispanic or Latino; Black refers to Black or African-American.

[c] 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 last two rows of Table 5-1 partition the Wave 4 Cohort members into their two distinct recruitment waves. The Wave 4.5 youth interview weighted response rate was 92.6 percent for those recruited in Wave 1 and 83.2 percent for those recruited in Wave 4. A possible explanation for this difference is that the Wave 1 sample members of the Wave 4 Cohort have persisted with the PATH Study over a number of years and therefore may be more cooperative than study participants recruited only 1 year earlier. In addition, all Wave 4 shadow youth in the Wave 4 Cohort were recruited at Wave 4.

Variation in response rates by subgroups is to be expected in large-scale data collection efforts. Apart from the lower Wave 4.5 response rate among Wave 4 shadow youth, none of the differences among the demographic and tobacco-use subgroups in Table 5-1 causes serious concern about potential nonresponse bias in the Wave 4.5 youth interview estimates. Section 5.2 examines the effectiveness of the Wave 4.5 weighting adjustments in addressing the lower response among younger youth.

5.2 Comparison Between Wave 4.5 Respondents and Those Eligible for Wave 4.5 Interview (Among the Wave 4 Cohort)

This section compares weighted estimates between the Wave 4.5 youth interview respondents in the Wave 4 Cohort and the Wave 4 respondents from the Wave 4 Cohort who were eligible for the Wave 4.5 youth interview. The weighted estimates cover demographic characteristics (shown in Table 5-2) and tobacco-use status (shown in Table 5-3). The structure of these tables is similar to the tables in Section 4.2.

Table 5-2. Comparison of Wave 4 demographic characteristics between Wave 4 respondents who were eligible for Wave 4.5 youth interview and Wave 4.5 youth interview respondents (Wave 4 Cohort)

Wave 4 characteristic	Wave 4 respondents who were eligible for Wave 4.5 youth interview		Wave 4 respondents who completed Wave 4.5 youth interview				
				Before Wave 4.5 weighting adjustment		After Wave 4.5 weighting adjustment	
	Unweighted count	A: Weighted percentage, using Wave 4 cross-sectional weight [95% confidence interval]	Unweighted count	B: Weighted percentage, using Wave 4 cross-sectional weight [95% confidence interval]	Difference in weighted percentages [B – A] [95% confidence interval]	C: Weighted percentage, using Wave 4.5 single-wave weight [95% confidence interval]	Difference in weighted percentages [C – A] [95% confidence interval]
Sex							
Male	7,479	51.1% [50.3%, 51.9%]	6,707	51.2% [50.3%, 52.0%]	0.1% [-0.2%, 0.4%]	51.1% [50.2%, 52.0%]	0.0% [-0.0%, 0.0%]
Female	6,892	48.9% [48.1%, 49.7%]	6,170	48.8% [48.0%, 49.7%]	-0.1% [-0.4%, 0.2%]	48.9% [48.0%, 49.8%]	0.0% [-0.0%, 0.0%]
Age group							
Under 12	2,073	16.5% [15.9%, 17.2%]	1,660	14.6% [14.1%, 15.3%]	-1.9% [-2.2%, -1.6%]	16.5% [15.9%, 17.2%]	0.0% [-0.0%, -0.0%]
12-13	4,754	33.3% [32.5%, 34.0%]	4,373	34.3% [33.5%, 35.1%]	1.0% [0.8%, 1.3%]	33.3% [32.5%, 34.1%]	0.0% [0.0%, 0.0%]
14-17	7,589	50.2% [49.4%, 51.0%]	6,885	51.0% [50.2%, 51.9%]	0.8% [0.5%, 1.2%]	50.2% [49.3%, 51.1%]	0.0% [0.0%, 0.0%]
Race/ethnicity[a]							
Hispanic	4,169	24.2% [23.5%, 24.9%]	3,769	24.3% [23.6%, 25.1%]	0.1% [-0.2%, 0.4%]	24.2% [23.5%, 25.0%]	0.0% [-0.0%, 0.1%]
Non-Hispanic White alone	6,516	52.4% [51.6%, 53.3%]	5,831	52.6% [51.7%, 53.4%]	0.1% [-0.2%, 0.4%]	52.4% [51.6%, 53.3%]	0.0% [-0.1%, 0.0%]
Non-Hispanic Black alone	1,813	13.2% [12.7%, 13.8%]	1,615	13.0% [12.5%, 13.6%]	-0.2% [-0.4%, 0.0%]	13.2% [12.6%, 13.8%]	0.0% [-0.1%, 0.0%]
Non-Hispanic other race or multiple races	1,377	10.2% [9.7%, 10.7%]	1,214	10.1% [9.6%, 10.6%]	-0.1% [-0.3%, 0.2%]	10.2% [9.6%, 10.7%]	0.0% [-0.0%, 0.1%]

Table 5-2. Comparison of Wave 4 demographic characteristics between Wave 4 respondents who were eligible for Wave 4.5 youth interview and Wave 4.5 youth interview respondents (Wave 4 Cohort) (continued)

Wave 4 characteristic	Wave 4 respondents who were eligible for Wave 4.5 youth interview		Wave 4 respondents who completed Wave 4.5 youth interview				
				Before Wave 4.5 weighting adjustment		After Wave 4.5 weighting adjustment	
	Unweighted count	A: Weighted percentage, using Wave 4 cross-sectional weight [95% confidence interval]	Unweighted count	B: Weighted percentage, using Wave 4 cross-sectional weight [95% confidence interval]	Difference in weighted percentages [B – A] [95% confidence interval]	C: Weighted percentage, using Wave 4.5 single-wave weight [95% confidence interval]	Difference in weighted percentages [C – A] [95% confidence interval]
Census region							
Northeast	2,038	16.2% [15.6%, 16.8%]	1,814	16.1% [15.5%, 16.8%]	0.0% [-0.3%, 0.2%]	16.2% [15.6%, 16.8%]	0.0% [-0.0%, -0.0%]
Midwest	3,203	21.4% [20.7%, 22.0%]	2,868	21.4% [20.7%, 22.1%]	0.1% [-0.2%, 0.3%]	21.4% [20.7%, 22.1%]	0.0% [0.0%, 0.0%]
South	5,432	38.5% [37.7%, 39.3%]	4,841	38.2% [37.4%, 39.1%]	-0.2% [-0.6%, 0.1%]	38.5% [37.6%, 39.3%]	0.0% [0.0%, 0.0%]
West	3,743	24.0% [23.3%, 24.7%]	3,395	24.2% [23.5%, 24.9%]	0.2% [-0.0%, 0.5%]	24.0% [23.2%, 24.7%]	0.0% [0.0%, 0.0%]

[a] Hispanic refers to Hispanic or Latino; Black refers to Black or African-American.

Table 5-3. Comparison of Wave 4 "ever tobacco-use" estimates between Wave 4 youth who were eligible for Wave 4.5 youth interview and Wave 4 youth who completed Wave 4.5 youth interview (Wave 4 Cohort)*

Wave 4 characteristic[a]	Wave 4 youth who were eligible for Wave 4.5 youth interview		Wave 4.5 youth interview respondents who completed Wave 4 youth interview				
				Before Wave 4.5 weighting adjustment		After Wave 4.5 weighting adjustment	
	Sample size	A: Weighted estimate, using Wave 4 cross-sectional weight [95% confidence interval]	Sample size	B: Weighted estimate, using Wave 4 cross-sectional weight [95% confidence interval]	Difference in weighted estimates [B – A] [95% confidence interval]	C: Weighted estimate, using Wave 4.5 single-wave weight [95% confidence interval]	Difference in weighted estimates [C – A] [95% confidence interval]
Overall	11,702	17.2% [16.4%, 18.0%]	10,694	17.0% [16.2%, 17.8]	-0.2% [-0.5%, 0.0%]	17.3% [16.5%, 18.1%]	0.1% [-0.0%, 0.3%]
Sex							
Male	6,032	18.8% [17.7%, 20.0%]	5,519	18.7% [17.6%, 19.8%]	-0.1% [-0.5%, 0.2%]	19.0% [17.9%, 20.2%]	0.2% [-0.0%, 0.5%]
Female	5,636	15.5% [14.5%, 16.6%]	5,142	15.2% [14.1%, 16.3%]	-0.4% [-0.8%, 0.0%]	15.5% [14.5%, 16.7%]	0.0% [-0.3%, 0.2%]
Age group[b]							
12-13	4,644	6.6% [5.8%, 7.5%]	4,275	6.6% [5.8%, 7.6%]	0.0% [-0.2%, 0.3%]	6.6% [5.8%, 7.6%]	0.0% [-0.2%, 0.2%]
14-17	7,058	24.5% [23.4%, 25.7%]	6,419	24.2% [23.1%, 25.4%]	-0.3% [-0.7%, 0.0%]	24.7% [23.6%, 25.9%]	0.2% [-0.0%, 0.4%]
Race/ethnicity[c]							
Hispanic	3,409	17.1% [15.8%, 18.4%]	3,154	17.0% [15.6%, 18.4%]	-0.1% [-0.5%, 0.3%]	17.4% [15.9%, 18.9%]	0.3% [-0.1%, 0.7%]
Non-Hispanic White alone	5,227	19.0% [17.8%, 20.3%]	4,754	18.9% [17.7%, 20.2%]	-0.1% [-0.5%, 0.3%]	19.2% [18.0%, 20.5%]	0.2% [-0.0%, 0.4%]
Non-Hispanic Black alone	1,492	13.0% [11.2%, 15.0%]	1,357	12.5% [10.6%, 14.6%]	-0.5% [-1.2%, 0.2%]	12.8% [10.9%, 15.0%]	-0.1% [-0.8%, 0.5%]
Non-Hispanic other race or multiple races	1,097	14.9% [12.8%, 17.2%]	992	14.3% [12.2%, 16.6%]	-0.6% [-1.4%, 0.3%]	14.8% [12.8%, 17.2%]	0.0% [-0.8%, 0.8%]

Table 5-3. Comparison of Wave 4 "ever tobacco-use" estimates between Wave 4 youth who were eligible for Wave 4.5 youth interview and Wave 4 youth who completed Wave 4.5 youth interview (Wave 4 Cohort) (continued)*

Wave 4 characteristic[a]	Wave 4 youth who were eligible for Wave 4.5 youth interview		Wave 4.5 youth interview respondents who completed Wave 4 youth interview				
				Before Wave 4.5 weighting adjustment		After Wave 4.5 weighting adjustment	
	Sample size	A: Weighted estimate, using Wave 4 cross-sectional weight [95% confidence interval]	Sample size	B: Weighted estimate, using Wave 4 cross-sectional weight [95% confidence interval]	Difference in weighted estimates [B – A] [95% confidence interval]	C: Weighted estimate, using Wave 4.5 single-wave weight [95% confidence interval]	Difference in weighted estimates [C – A] [95% confidence interval]
Census region							
Northeast	1,639	17.4% [15.6%, 19.3%]	1,495	17.2% [15.3%, 19.1%]	-0.2% [-0.8%, 0.4%]	17.3% [15.5%, 19.3%]	0.0% [-0.6%, 0.5%]
Midwest	2,590	18.0% [16.5%, 19.5%]	2,368	18.2% [16.7%, 19.9%]	0.3% [-0.1%, 0.6%]	18.6% [17.1%, 20.3%]	0.7% [0.3 %, 1.1%]
South	4,424	17.0% [15.9%, 18.2%]	4,013	16.6% [15.5%, 17.8%]	-0.4% [-0.9%, 0.1%]	17.0% [15.9%, 18.2%]	0.0% [-0.4%, 0.4%]
West	3,049	16.7% [14.5%, 19.1%]	2,818	16.2% [14.1%, 18.6%]	-0.5% [-0.9%, -0.0%]	16.6% [14.3%, 19.1%]	-0.1% [-0.5%, 0.3%]

*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.

[a] For each Wave 4 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 5-3 includes a subset of the youth in Table 5-2 because there are no Wave 4 interview data for the Wave 4.5 youth interview respondents who were shadow youth at Wave 4.

[c] Hispanic refers to Hispanic or Latino; Black refers to Black or African-American.

Table 5-2 compares Wave 4 demographic characteristics between the Wave 4.5 youth interview respondents and those eligible for the Wave 4.5 youth interview. The percentages sum to 100 percent[14] over the categories associated with each characteristic. For sex, race/ethnicity, and census region, the "before Wave 4.5 weighting adjustment" distributions among the Wave 4.5 youth interview respondents were almost the same as the distributions among those eligible for the Wave 4.5 youth interview. Similar to the result shown in Table 4-2 for the Wave 1 Cohort, the most noticeable underrepresentation among the Wave 4.5 youth respondents was the Wave 4 Cohort "under 12" age group.

For the "after Wave 4.5 weighting adjustment" differences, the point estimates were zero (to one decimal place) and the estimated confidence intervals either included zero or were in very close proximity to zero for each of the Wave 4 characteristics examined: sex, age, race/ethnicity, and census region.

Table 5-3 shows the comparison of Wave 4 "ever tobacco-use" rates between Wave 4 youth interview respondents who completed the Wave 4.5 youth interview and all the Wave 4 youth interview respondents eligible for the Wave 4.5 youth interview (among the Wave 4 Cohort). (Table 5-3 includes a subset of the youth in Table 5-2 because there are no Wave 4 interview data for the Wave 4.5 youth interview respondents who were shadow youth at Wave 4.) For both the "before Wave 4.5 weighting adjustment" and the "after Wave 4.5 weighting adjustment" estimates, differences between respondents and those eligible for the interview were no more than 0.7 percent. The largest difference was for the Wave 4 "ever tobacco-use" measure among youth from the Midwest census region after Wave 4.5 weighting adjustment. However, most of the confidence intervals either included or were in the proximity of zero.

Overall, these results reflect the use of both demographic variables and tobacco-use measures from Wave 4 for calibrating the Wave 4.5 single-wave weight for the Wave 4 Cohort. Assuming that the Wave 4 demographic and tobacco-use characteristics in Tables 5-2 and 5-3 are correlated with key tobacco- and health-related outcome measures in Wave 4.5 of the PATH Study, these results indicate little if any nonresponse bias in the youth interview estimates for the Wave 4 Cohort due to attrition from Wave 4 to Wave 4.5.

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

5.3 Comparison of Youth Cigarette-Smoking Estimates to Other National Studies

Table 5-4 compares estimates for two common measures of youth cigarette-smoking based on the Wave 4.5 youth interview respondents from the Wave 4 Cohort of the PATH Study to similar estimates based on data from NHANES 2015-2016, NSDUH 2017, and NYTS 2018. As mentioned in Section 3.2.3, the data for all three external studies were restricted to ages 12-17 to match the PATH Study definition of youth. The primary measure of cigarette smoking among youth is whether the youth has ever tried smoking a cigarette; even one or two puffs. Another measure is current smoking, defined as having smoked at all in the past 30 days. (See Appendix A for details.)

PATH Study estimates, 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 4.5 single-wave weight for the Wave 4 Cohort. 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. Comparison of youth cigarette-smoking estimates from PATH Study Wave 4.5 youth interview (Wave 4 Cohort) and other national studies*

Smoking measure and subgroup[a]	PATH Study Wave 4.5 (2017-2018) sample size	Percentage from PATH Study Wave 4.5 (2017-2018) [95% confidence interval]	Percentage from NHANES 2015-2016 [95% confidence interval]	Percentage from NSDUH 2017 [95% confidence interval]	Percentage from NYTS 2018 [95% confidence interval]
Ever tried cigarette smoking, even one or two puffs, overall	12,732	8.8% [8.3%, 9.4%]	12.9% [10.4%, 15.9%]	10.7% [10.0%, 11.4%]	13.7% [12.6%, 15.0%]
Ever tried smoking, male	6,604	9.5% [8.7%, 10.3%]	14.6% [10.1%, 20.6%]	11.2% [10.1%, 12.3%]	16.7% [15.1%, 18.3%]
Ever tried smoking, female	6,083	8.1% [7.4%, 9.0%]	11.2% [8.6%, 14.4%]	10.2% [9.2%, 11.2%]	15.4% [13.8%, 17.2%]
Ever tried smoking, age 12-13	3,770	2.4% [1.9%, 3.1%]	2.8% [1.1%, 6.7%]	3.0% [2.4%, 3.8%]	7.0% [5.9%, 8.3%]
Ever tried smoking, age 14-17	8,962	12.1% [11.4%, 12.9%]	17.5% [14.2%, 21.5%]	14.1% [13.2%, 15.1%]	20.6% [19.0%, 22.2%]
Have smoked in past 30 days, overall	12,900	2.7% [2.4%, 3.0%]	3.3% [2.0%, 5.2%]	3.2% [2.9%, 3.6%]	4.1% [3.6%, 4.7%]

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

[a] Unless otherwise noted, the subgroup is for youth ages 12 to 17.

The PATH Study point estimates were lower than the estimates from the other studies. However, the confidence intervals overlapped or almost overlapped between Wave 4.5 of the PATH Study, NHANES 2015-2016, and NSDUH 2017 for four of the six estimates presented in Table 5-4, with ever tried smoking, overall and for ages 14-17, being the exceptions. None of the confidence intervals overlapped between Wave 4.5 of the PATH Study and NYTS 2018 for the estimates of cigarette smoking in Table 5-4.

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 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 mode of administration, context, question order and wording, and year of data collection. The PATH Study, NHANES, and NSDUH use audio computer-assisted self-interview (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. NYRBS is a survey of only high school students. 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). Consistent with this research and the results in Table 5-4, the NYRBS 2017 estimated that 26.9 percent of students ages 14-17 have ever tried cigarette smoking and the confidence interval for this estimate does not overlap with the confidence interval for the corresponding PATH Study estimate, which is lower.[15]

The contexts and purposes of these surveys also differ: NHANES is a general health survey 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.

[15]Cigarette-smoking estimates for NYRBS 2017 were not included in Table 5-4 because only the estimate for ages 14-17 can be computed.

Finally, according to Monitoring the Future (Johnston et al., 2019), cigarette smoking among 8th to 12th graders dropped from 2015 to 2018 (the percentage of youth who had ever tried smoking dropped by 4.2 percentage points among 8th graders, 3.9 percentage points among 10th graders, and 7.3 percentage points among 12th graders). The lower percentages found in the PATH Study may reflect, in part, a continuation of this trend; however, cigarette-smoking rates have leveled off among 8th and 10th grade students since 2017.

5.4 Comparison of Youth ENDS-Use Estimates to Other National Studies

Table 5-5 compares estimates for youth ever use and current use of ENDS based on the Wave 4.5 youth interview respondents from the Wave 4 Cohort of the PATH Study to similar estimates based on data from NYTS 2018[16] and NYRBS 2017. The age ranges of the PATH Study respondents at Wave 4.5 and the NYRBS 2017 respondents do not fully overlap. NYRBS is a survey of only high school students and has very few respondents under age 14; Wave 4.5 respondents could be no older than 17. Therefore, the PATH Study, NYTS, and NYRBS estimates were restricted to ages 14-17 for comparison. The primary measure of ENDS use among youth is whether the youth has ever tried using ENDS. A second measure is current ENDS use, defined as having used ENDS at all in the past 30 days. (See Appendix B for details.)

PATH study estimates and 95 percent confidence intervals are shown for 14- to 17-year-olds as a whole, and by sex. The point estimates for the PATH Study were calculated using the Wave 4.5 single-wave weight for the Wave 4 Cohort and confidence intervals required the corresponding replicate weights for variance estimation. Point estimates and 95 percent confidence intervals are reported for NYTS 2018 and NYRBS 2017 as well.

[16] NYTS 2018 asks about e-cigarette use and not ENDS use. However, the definition of e-cigarettes provided to NYTS respondents just prior to asking about e-cigarette use frequency is virtually identical to the definition of "electronic nicotine products" provided to PATH Study respondents. Specifically, the PATH Study includes e-cigarettes as a specific kind of "electronic nicotine product," whereas the NYTS defines e-cigarettes as encompassing all ENDS use (e.g., e-cigarette, e-cigar, e-hookah, e-pipe). With these definitions in mind, NYTS 2018 e-cigarette use estimates are equivalent to ENDS use estimates from Wave 4.5 of the PATH Study and NYRBS 2017 in Table 5.5.

Table 5-5. Comparison of ENDS-use estimates from PATH Study Wave 4.5 youth interview (Wave 4 Cohort) and other national studies*

ENDS-use measure and subgroup	PATH Study Wave 4.5 (2017-2018) sample size	Percentage from PATH Study Wave 4.5 (2017-2018) [95% confidence interval]	Percentage from NYTS 2018 [95% confidence interval]	Percentage from NYRBS 2017 [95% confidence interval]
Ever used ENDS, age 14-17	8,949	23.5% [22.5%, 24.4%]	32.5% [30.4%, 34.6%]	40.4% [37.5%, 43.4%]
Ever used ENDS, male age 14-17	4,613	23.9% [22.6%, 25.3%]	33.8% [31.7%, 36.0%]	42.6% [40.2%, 45.0%]
Ever used ENDS, female age 14-17	4,312	23.0% [21.6%, 24.4%]	31.1% [28.6%, 33.7%]	38.4% [34.3%, 42.6%]
Have used ENDS in past 30 days, age 14-17	9,118	9.7% [9.1%, 10.4%]	17.6% [16.0%, 19.4%]	11.8% [10.0%, 13.9%]

*Estimates from the 2018 NYTS and 2017 NYRBS were obtained using public use files for those surveys.

The PATH Study point estimates were significantly lower than were those from NYTS 2018 and NYRBS 2017. Only the estimate of ENDS use in the past 30 days among ages 14-17 has an overlap in the confidence intervals. As discussed in Section 5.3, some differences in the estimates may be attributable to differences in context, question order and wording, mode of administration, timeframes of the surveys, and measurement error. In particular, NYTS and NYRBS are self-administered in the school whereas the PATH Study generally is administered in the home. Section 5.3 notes that school-based surveys tend to produce higher smoking rates and it is possible that this also apply to estimates of ENDS use. It is uncertain which factors are the key drivers of the differences in ENDS-use estimates observed between the PATH Study and other national studies.

6. Summary of Findings

This chapter summarizes the PATH Study's Wave 4.5 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 4.5 youth interview response rates.
- The national representativeness of Wave 4.5 youth respondents and statistical weighting adjustments to reduce potential nonresponse bias.

6.1 Wave 1 Cohort

The Wave 1 Cohort unweighted response rate was similar to the weighted response rate for the Wave 4.5 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 74.6 percent.
- Moderate differences (i.e., of no more than 6 percent) were found by Wave 1 sex, race/ethnicity, and census region.
- The weighted response rate for those under age 12 was noticeably lower than for those ages 12-13 at Wave 1. This is likely because the persons in the youngest age group were shadow youth at Wave 1 and so have had less direct involvement with the study than youth have been interviewed since Wave 1. (A similar response rate pattern was observed in Waves 2, 3, and 4.)
- The weighted response rate among Wave 1 youth was 4.2 percentage points 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 4.5 youth estimates for the Wave 1 Cohort, as variation in response rates by subgroups is to be expected in large-scale data collection efforts and weighting adjustments are used to address nonresponse.

As a result of comparing select demographic and tobacco-use characteristics between the Wave 4.5 youth interview respondents and those eligible for the Wave 4.5 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 4.5 longitudinal weighting adjustments, the underrepresentation of Wave 1 shadow youth was essentially eliminated. Estimates of Wave 1 "ever tobacco use" among those who were youth at Wave 1 remained slightly lower for the Wave 4.5 respondents than for those eligible for the youth interview overall, and were slightly higher for non-Hispanic Black alone youth.

6.2 Wave 4 Cohort

The Wave 4 Cohort unweighted response rate was similar to the weighted response rate for the Wave 4.5 youth interview (conditioning on Wave 4 response) overall and by subgroups, although there were differences in weighted response rates across subgroups.

- The weighted overall response rate was 89.1 percent.
- The weighted response rates varied by 13.0 percentage points for Wave 4 age and 9.4 percentage points for recruitment wave. Response rates were lower in the "under 12" age group and among youth recruited in Wave 4 – these are highly correlated characteristics because all Wave 4 shadow youth in the Wave 4 Cohort were from the Wave 4 replenishment sample. Wave 4.5 was the first time that such shadow youth were asked to complete a PATH Study interview.

These differences in response rates do not necessarily indicate nonresponse bias in the Wave 4.5 youth estimates for the Wave 4 Cohort, as variation in response rates by subgroups is to be expected in large-scale data collection efforts and weighting adjustments are used to address nonresponse.

As a result of comparing select demographic and tobacco-use characteristics between the Wave 4.5 youth interview respondents and those eligible for the Wave 4.5 youth interview, the notable underrepresentation due to panel attrition among the Wave 4 Cohort was for the Wave 4 "under 12" age group, i.e., Wave 4 shadow youth.

After Wave 4.5 longitudinal weighting adjustments, the underrepresentation of Wave 4 shadow youth was essentially eliminated. Estimates of Wave 4 "ever tobacco use" among those youth in the Midwest census region at Wave 4 remained slightly higher for the Wave 4.5 respondents than for those eligible for the youth interview.

The PATH Study estimates of youth cigarette smoking and ENDS use among 14- to 17-year-olds were lower than the estimates from the other national studies. The confidence intervals for cigarette smoking overlapped between Wave 4.5 of the PATH Study, NHANES 2015-2016, and NSDUH

2017 for most estimates. However, the confidence intervals for ENDS use mostly did not overlap between the PATH Study and NYTS 2018 or NYRBS 2017. It is uncertain whether the differences are due to nonresponse bias in one or more of the studies compared or a variety of other differences between the studies.

6.3 General Conclusions

Assuming that the demographic 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 4.5 of the PATH Study. However, the Wave 1 Cohort analyses hint at the challenges faced by cohort studies as attrition accumulates and confirm the timeliness of the introduction of the Wave 4 replenishment sample to the study.

The findings for the both the Wave 1 Cohort and Wave 4 Cohort suggest that the longitudinal weighting adjustments satisfactorily address the potential bias due to attrition. However, at Wave 4.5, the Wave 4 Cohort produces cigarette-smoking and ENDS-use estimates that are lower than those from other national studies. It is difficult to determine the primary reason(s) for these results due to the many differences between the PATH Study and the surveys compared.

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, 2017, https://www.datafiles.samhsa.gov/study-dataset/national-survey-drug-use-and-health-2017-nsduh-2017-ds0001-nid17939.

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 Youth Risk Behavior Survey Data.* Atlanta, GA: Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Division of Adolescent and School Health, 2017, https://www.cdc.gov/healthyyouth/data/yrbs/data.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, 2018, 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.

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. (2019). *Monitoring the Future national survey results on drug use: 1975-2018: 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-overview2018.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.

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

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

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.

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 define youth cigarette smoking in the PATH Study, NHANES, NSDUH, and NYTS, and describes the populations included in the estimates from those surveys.

Note that although the questions used to define cigarette smoking are similar among the surveys, small differences could have an effect on the answers given. In the PATH Study, the youth interview 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."

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

Table A-1. Questions used to define youth cigarette smoking in the PATH Study, NHANES, NSDUH, and NYTS

PATH Study	NHANES	NSDUH	NYTS
Questions 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 2–4.5 new baseline youth question, yes) and "In the past 12 months, have you smoked a cigarette, even one or two puffs?" (Wave 2–4.5 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 2–4.5 new baseline youth question, yes) and "In the past 12 months, have you smoked a cigarette, even one or two puffs?" (Wave 2–4.5 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-1. 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. The target population for the Wave 4 Cohort at Wave 4.5 was the Wave 4 target population residing in the U.S. and 12 to 17 years old at Wave 4.5, except for those who were incarcerated at that time. Thus, it includes Wave 4 respondents who were on active duty or living in a health care institution (e.g., a mental health hospital) but not those in a correctional facility at Wave 4.5.	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-1. 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
ENDS-Use Questions in the PATH Study and Other Surveys

Table B-1 lists the questions used to ask about ever and current ENDS use of Wave 4.5 youth respondents in the PATH Study and in the surveys used for comparison, and describes the populations included in the estimates from those surveys.

Administration of the questions differs between the PATH Study, NYRBS, and NYTS. In the PATH Study, the questions generally are asked in the home where the parent or guardian may be present. Questions in NYRBS and NYTS are self-administered by students at their school.

Table B-1. Questions used to define youth ever and current ENDS use in the PATH Study, NYTS, and NYRBS

PATH Study	NYTS	NYRBS
Questions to define ever ENDS use (answers defining ever ENDS use given in parentheses)		
"Have you ever used an e-cigarette, such as NJOY, Blu, or Smoking Everywhere, even one or two times?" (Wave 1 youth, yes) and "Have you ever used an electronic nicotine product, even one or two times?" (Wave 2 youth who had not used an e-cigarette at Wave 1; Wave 2 – 4.5 new baseline youth question, yes) and "In the past 12 months, have you used an electronic nicotine product, even one or two times?" (Wave 2 youth who had used an e-cigarette at Wave 1; Wave 3 – 4.5 continuing youth, yes)	"Have you ever used an e-cigarette, even once or twice?" (yes)	"Have you ever used an electronic vapor product?" (yes)
Question to define current ENDS use (answers defining current ENDS use given in parentheses)		
"In the past 30 days, on how many days did you use an electronic nicotine product?" (1 or more days)		"During the past 30 days, on how many days did you use an electronic vapor product?" (1 or more days)
Age range included in estimates		
14-17	14-17	14-17

Table B-1. Questions used to define youth ever and current ENDS use in the PATH Study, NYTS, and NYRBS (continued)

PATH Study	NYTS	NYRBS
Exclusions from population		
The Wave 4 Cohort target population included only the U.S. civilian, noninstitutionalized population. The target population for the Wave 4 Cohort at Wave 4.5 was the Wave 4 target population residing in the U.S. and 12 to 17 years old at Wave 4.5, except for those who were incarcerated at that time. Thus, it includes Wave 4 respondents who were on active duty or living in a health care institution (e.g., a mental health hospital) but not those in a correctional facility at Wave 4.5.	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.	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.
Other comments		
	Self-administered survey in classroom.	Self-administered survey in classroom.