

Attachment 21
PATH Study Interim Report

Population Assessment of Tobacco and Health Study

Interim Report to the Office of
Management and Budget on Wave
1 and Wave 2 Data and
Biospecimen Collection

May 18, 2015

Submitted by:

Kevin P. Conway, Ph.D.

Deputy Director

Division of Epidemiology, Services, and Prevention Research

National Institute on Drug Abuse

6001 Executive Blvd., Room 5185

Rockville, MD 20852

Phone: 301-443-8755

Table of Contents

<u>Section</u>		<u>Page</u>
1	Introduction	1
1.1	Purpose of Interim Report (Terms of Clearance).....	1
1.2	Sample Design.....	2
1.2.1	Overview of Sample Design for Wave 1 (Baseline Wave).....	2
1.2.2	Overview of Sample Design for Wave 2.....	3
2	Wave 1.....	5
2.1	Response Rates.....	5
2.1.1	Household Screener	5
2.1.2	Extended Interviews	8
2.1.3	Biospecimen Collections	12
2.2	Nonresponse Bias Analysis	15
2.2.1	Method	15
2.2.2	Results	18
2.3	Statistical Approach for Addressing Nonresponse.....	30
2.3.1	Interviews.....	30
2.3.2	Biospecimens.....	41
2.4	Summary of Findings	45
3	Wave 2.....	48
3.1	Response Rates.....	48
3.1.1	Retention Rates for Continuing Adults and Continuing Youth.....	48
3.1.2	Recruitment Rates for Aged-up Adults and Aged-up Youth	52
3.1.3	Biospecimen Collections	58

<u>Section</u>	<u>Page</u>
3.2 Nonresponse Bias Analysis	60
3.3 Statistical Approach for Addressing Nonresponse	65
3.4 Summary of Findings	66
4 Conclusions and Implications for the Study Going Forward	69
References	72

Appendix

A Cigarette Smoking Questions on the PATH Study and Other Surveys	A-1
---	-----

Table

2-1 PATH Study Wave 1 response rates by address characteristics: Household Screener	8
2-2 PATH Study Wave 1 response rates by respondent characteristics: Adult Extended Interview	11
2-3 PATH Study Wave 1 response rates by respondent characteristics: Youth Interview	12
2-4 PATH Study Wave 1 response rates by respondent characteristics: Biospecimen collections	14
2-5 Race by age distribution, based on the household enumeration: IPS weights	19
2-6 Distribution of male and female adults listed in the household enumeration: IPS weights	19
2-7 Distribution of household size based on households responding to the Household Screener: IPS weights	20
2-8 Distribution of number of adults based on households responding to the Household Screener: IPS weights	20

<u>Table</u>	<u>Page</u>
2-9	Distribution of number of youth ages 12-17 based on households responding to the Household Screener: IPS weights 20
2-10	Demographic distributions based on adults responding to the Adult Extended Interview, and based on adults providing urine or blood specimens: IPS weights 21
2-11	Comparison of education level and health insurance status based on adults completing the Adult Extended Interview and based on adults providing urine or blood specimens: IPS weights..... 22
2-12	Current cigarette smoking based on adults responding to the Adult Extended Interview: IPS weights 26
2-13	Current cigarette smoking based on adults providing biospecimens in Wave 1: IPS weights 27
2-14	Demographic distributions based on youth ages 12-17 who completed the Youth Interview: IPS weights..... 28
2-15	Cigarette smoking* based on youth ages 12-17 who completed the Youth Interview: IPS weights 29
2-16	Race by age distribution, based on household enumeration..... 34
2-17	Distribution of male and female adults listed in the household enumeration..... 34
2-18	Distribution of household size based on households responding to the Household Screener..... 35
2-19	Distribution of number of adults based on households responding to the Household Screener..... 35
2-20	Distribution of number of youth ages 12-17 based on households responding to the Household Screener..... 35
2-21	Demographic distributions based on adults responding to the Adult Extended Interview 36

<u>Table</u>	<u>Page</u>
2-22	Comparison of education level and health insurance status based on adults responding to the Adult Extended Interview 37
2-23	Current cigarette smoking based on adults responding to the Adult Extended Interview 38
2-24	Demographic distributions based on youth ages 12-17 who completed the Youth Interview 39
2-25	Cigarette smoking* based on youth ages 12-17 who completed the Youth Interview 40
2-26	Demographic distributions based on adults from whom urine or blood specimens were collected 42
2-27	Comparison of education level and health insurance status based on adults from whom urine or blood specimens were collected 43
2-28	Current cigarette smoking based on adults from whom biospecimens were collected 44
2-29	Summary of PATH Study Wave 1 overall response rates 45
3-1	Status of Wave 2 released cases from replicate group 1, as of April 22, 2015 49
3-2	PATH Study Wave 2 predicted retention rates by Wave 1 characteristics: Adult Interview (continuing adults) 53
3-3	PATH Study Wave 2 predicted retention rates by respondent characteristics: Youth Interview (continuing youth) 54
3-4	PATH Study Wave 2 predicted recruitment rates by respondent characteristics: Adult Interview (aged-up adults) 56
3-5	PATH Study Wave 2 predicted recruitment rates by respondent characteristics: Youth Interview (aged-up youth) 57
3-6	PATH Study Wave 2 response rates by respondent characteristics: Urine collection (continuing adults) 59

<u>Table</u>		<u>Page</u>
3-7	PATH Study Wave 2 response rates by respondent characteristics: Biospecimen collections (aged-up adults)	60
3-8	Comparison of Wave 2 Continuing Adult Interview respondents with finalized and provisional nonrespondents	62
3-9	Comparison of Wave 1 tobacco use rates for Wave 2 Continuing Adult Interview respondents with finalized and provisional nonrespondents.....	63
3-10	Comparison of Wave 2 Aged-up Adult and Continuing Youth Interview respondents with provisional nonrespondents	64
3-11	Summary of PATH Study predicted response rates for Wave 2.....	67
A-1	Question used to define “current smoking” in the PATH Study, TUS-CPS, NHIS, NHANES, and NSDUH	2
A-2	Questions used for youth cigarette smoking in the PATH Study, NHANES, NSDUH, and NYTS	3

Introduction

1

The National Institute on Drug Abuse (NIDA), in partnership with the Food and Drug Administration (FDA), has prepared this report in response to the terms of clearance by the Office of Management and Budget (OMB) in its approval of Wave 2 of the Population Assessment of Tobacco and Health (PATH) Study.

1.1 Purpose of Interim Report (Terms of Clearance)

OMB approved the PATH Study's Revision Request for Wave 2 on September 8, 2014 (0925-0664). The terms of clearance of OMB's approval state: "Before submitting the information collection request for Wave 3 to OMB, NIDA/FDA should report to OMB regarding: a) the response rates associated with the full baseline wave, including screening, interview completion, and bio-specimen response; b) Wave 2 retention, recruitment rates for the 'age in to adult' and 'age in of shadow' subsamples; c) the results of nonresponse analysis and statistical approach for addressing non-response, as well as implications for the study going forward; and d) the statistical approach to be applied to the bio-specimen data to address potential non-response bias from lower consent and cooperation rates with this aspect of the study."

This report is organized in sections that correspond to OMB's terms of clearance: Section 2 presents the Wave 1 response rates, results of a nonresponse analysis, and statistical approach for addressing nonresponse. Section 3 presents the Wave 2 response rates (retention and recruitment rates), results of a nonresponse analysis, and statistical approach for addressing nonresponse. Section 4 summarizes the findings and considers their implications.

The report covers the PATH Study's Wave 1 and data collected from a probability subsample of Wave 2 between October 23, 2014 and April 22, 2015. Response rates for Wave 1 and Wave 2 are compared throughout this report with corresponding rates projected in the PATH Study's respective non-substantive change request to OMB for Wave 1 and revision request to OMB for Wave 2. For Wave 1, the comparisons refer to the best-case and worst-case scenarios for the entire sample, provided in "Attachment 22." ("Attachment 22" is part of Supporting Statement B of the PATH Study's non-substantive change request for Wave 1.)

1.2 Sample Design

This section provides an overview of the sample design for the PATH Study and a description of replicate group 1 on which the interim report results are based for Wave 2. Information on the study background and overall design is provided in Supporting Statement A of the PATH Study's Revision Request for Wave 2.

1.2.1 Overview of Sample Design for Wave 1 (Baseline Wave)

The target population of the PATH Study is the civilian, non-institutionalized U.S. population (excluding Puerto Rico) 12 years of age and older. Active duty military personnel and residents of group quarters are excluded, with the exception of college students. For Wave 1, a four-stage stratified area probability sample design was used, with a two-phase design for sampling the adult cohort 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 were formed and then a sample of these segments was drawn. At the third stage, a sample of addresses was drawn from a sampling frame that consists of the residential addresses in the U.S.; the main source of these addresses was the Postal Service (USPS) Computerized Delivery Sequence Files (CDSFs).

The fourth stage selected persons from the sampled households. A roster of all the members in the sampled household was constructed using the Household Screener. The Household Screener was used to ask one adult household member (referred to as the household informant) to list members of the household and provide demographic, and, for adults, tobacco use information about each for use in sampling three main groups of interest:

- Adults (up to two adults per household);
- Children ages 12 to 17 (referred to as “youth,” generally up to two per household); and
- Children ages 9 to 11 (referred to as “shadow youth,” generally up to two per household) to be enrolled in the youth cohort in later waves of the study on reaching 12 years of age.

Two-phase sampling was used for adult selection due to potential misreporting of the tobacco use status of other adult members of the household by the household informant. The Phase 1 sampling

depended on the age, race, and tobacco use information provided by the household informant. The Phase 2 sampling was based on the sampled individual's self-reported age, race, and tobacco use status, obtained by interviewing the individuals sampled at Phase 1. The sampling rates for the two phases were designed to achieve large enough sample sizes for young adults (ages 18 to 24) and adult tobacco users of all ages. The tobacco use status reported by the household informant is referred to as “Phase 1 tobacco use status”; the self-reported tobacco use information obtained during Phase 2 screening is referred to as “Phase 2 tobacco use status.”

Because the full sample was selected using probability sampling methods, it is representative of the U.S. civilian non-institutionalized population 12 years of age and older. The PATH Study Wave 1 sample was divided into four replicate groups, consisting of probability samples of approximately 20 percent, 30 percent, 30 percent, and 20 percent of the sampled segments, respectively, within each sampled PSU. Each separate replicate group was a probability sample from the set of segments in the frame and therefore also representative of the civilian non-institutionalized U.S. population. The replicate groups were released to the field in a sequential manner (replicate group 1 in September 2013, replicate group 2 in November 2013, replicate group 3 in February and March 2014, and replicate group 4 in May 2014).

1.2.2 Overview of Sample Design for Wave 2

The PATH Study is currently conducting Wave 2, which is the first follow-up wave for participants in Wave 1. In addition, youth from the shadow youth sample established at Wave 1 who turn age 12 by Wave 2, and who are permitted by a parent or guardian to participate in the study, are asked for assent to be interviewed for the first time at Wave 2. Similarly, persons in the youth sample at Wave 1 who reach age 18 by Wave 2 are asked to complete a Wave 1 version of the adult instrument, and to provide urine and blood samples for the first time.

For Wave 2, the PATH Study is subsampling approximately 14,500 adults for urine collection from adults who provided urine at Wave 1. Among these adults, at least 10,000 are expected to provide a urine sample again at Wave 2. The Study is also collecting urine and blood samples from consenting aged-up adults.

Results in this Interim Report for Wave 2 are based on data collected about persons in replicate group 1 as of the cutoff date for the analyses on April 22, 2015. Wave 2 cases are released for follow up in monthly groups with the goal of completing the Wave 2 interview as close as possible to the

one-year anniversary date of the Wave 1 interview. Cases are fielded approximately one month prior to the anniversary date. If more than one Wave 1 respondent is from the same household, the cases from that household are clustered for simultaneous release based on the date of the first completed interview for a sampled adult or youth in the household. Thus, while the Wave 2 sample release is not explicitly tied to the release of the Wave 1 replicate groups, it does correspond with the distribution of completion dates for Wave 1 interviews. That is, a high percentage of cases in the first Wave 2 release groups came from replicate group 1, and replicate group 1 has higher percentages of released and finalized cases than the other replicate groups.

Wave 1 **2**

The PATH Study completed the Wave 1 data and biospecimen collection in December 2014.¹ This section presents findings on the response rates for Wave 1, on the nonresponse analysis, and on the Study's statistical approach for addressing nonresponse.

2.1 Response Rates

This section summarizes the Wave 1 response rates for the Household Screener, Adult and Youth Extended Interviews, and biospecimen collections.

2.1.1 Household Screener

The Household Screener combined typical screener functions (e.g., enumerating the household, collecting demographic information about each member and some household-level data, and selecting participants for the study) with a special purpose for the PATH Study, collecting basic information on each adult's tobacco use.² This allowed the Study to classify the adult with sufficient validity for potential selection as a participant based on the PATH Study's sampling strata on tobacco use and demographic characteristics. Field interviewers conducted the Household Screener in person using computer-assisted personal interviewing (CAPI). The Study released 167,525 addresses; of these, 141,509 Household Screener cases were finalized (i.e., were completed or were finalized nonrespondents) during Wave 1.

¹ A small number of blood collections for Wave 1 were completed in January 2015.

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 at the time when Wave 1 was designed, such as electronic cigarettes or e-cigarettes.

Method

The PATH Study computed the response rates presented in this report in accordance with the response rate formula provided by OMB in its “Standards and Guidelines for Statistical Surveys” (2006). This formula calls for calculating unweighted unit response rates (RRU) as the ratio of the number of completed cases (or sufficient partials) to the number of in-scope sample cases.³ The different categories of cases that comprise the total number of in-scope cases are defined as follows:

- C = number of completed cases or sufficient partials;
- R = number of refused cases;
- NC = number of noncontacted sample units known to be eligible;
- O = number of eligible sample units not responding for reasons other than refusal;
- U = number of sample units of unknown eligibility, not completed; and
- e = estimated proportion of sample units of unknown eligibility that are eligible.

The unweighted unit response rate represents a composite of these components:

$$RRU=C/(C+R+NC+O+e(U))$$

In computing the response rates, refused cases, noncontacted sample units known to be eligible, and eligible sample units not responding for reasons other than refusal were combined into nonresponse cases (NR):

$$RRU=C/(C+NR+e(U))$$

The response rates were weighted to compensate for unequal probabilities of selection due to planned oversampling of individuals with certain characteristics (i.e., young adults, African-American adults, and adult tobacco users). Although the unweighted response rate measures operational success, the weighted response rate is needed to generalize to the population. The Household Screener inverse probability of selection (IPS) weights were calculated as the inverse of the selection probabilities for the sampled households. (See Section 2.2 for additional information on weighting). Note that the weights used for this report are preliminary; estimates calculated using the final versions of the weights may differ slightly from the estimates presented in this report.

³ The sample does not have any partial completes.

Table 2-1 provides overall response rates for the Household Screener and response rates for subgroups of sampled households that belong to Census block groups with various characteristics. In addition to the overall row, the table includes rows on education, race, ethnicity, and poverty status subgroups. For example, the weighted response rate for addresses in Census block groups with “high” levels of education (>29.6% of persons ages 25 and older with Bachelor’s degrees) was 48.3 percent; it was 58.0 percent for addresses in Census block groups with “low” levels of education. Comparing subgroups of responding and nonresponding households informed an assessment of the extent to which the responding addresses represented all sampled addresses and, ultimately, the population of inference. For this purpose, subgroups were defined by the characteristics of the Census block groups in which the sampled addresses were located using information from the 5-year (2009 to 2013) American Community Survey (ACS).⁴ The “high” and “low” subgroup categories were defined relative to the nationwide percentage of persons having the characteristic: block groups with percentages below the national average for the characteristic were classified as low, and those with percentages above the national average were classified as high.

Results

As indicated in Table 2-1, the weighted overall response rate for the Household Screener was 54.0 percent. The weighted response rates for subgroups defined by neighborhood characteristics indicate the subgroups differed from one another by as much as 9.7 percentage points. The differences among subgroups on weighted response rates were 9.7 percentage points for education, 2.7 percentage points for race, 1.1 percentage points for ethnicity, and 7.2 percentage points for poverty status. The overall response rate for the Household Screener was lower than the projected rate of 70 percent previously presented to OMB in Attachment 22, but it exceeded the worst-case scenario response rate of 39.7 percent.

⁴ Information from the 5-year (2009 to 2013) rather than the 1-year (2013) ACS was used because 1-year ACS estimates are not provided for smaller geographies such as Census tracts or block groups. The 5-year ACS estimates, which are based on the accumulated sample from 2009 to 2013, are the only estimates from ACS that provide information at the tract level and for smaller geographies (see http://www.census.gov/acs/www/guidance_for_data_users/estimates/).

Table 2-1. PATH Study Wave 1 response rates by address characteristics: Household Screener

Neighborhood characteristic ^a	A: Completed (n)	B: Finalized nonresponse ^b (n)	C: Unknown eligible estimated to be eligible ^c (n)	Unweighted response rate for Wave 1 ^d (%)	Weighted response rate for Wave 1 (%)
Overall	79,198	62,332	4,760	54.1	54.0
Education (% with Bachelor's degree)					
High > 29.6%	29,817	28,380	3,194	48.6	48.3
Low <= 29.6%	49,381	33,952	1,566	58.2	58.0
Race (% Black alone or in combination)					
High > 13.8%	20,903	15,200	1,144	56.1	56.0
Low <= 13.8%	58,295	47,132	3,616	53.5	53.3
Ethnicity (% Hispanic)					
High > 17.1%	21,596	16,522	1,186	54.9	54.8
Low <= 17.1%	57,602	45,810	3,574	53.8	53.7
Poverty Status					
High > 15.8%	31,085	20,104	1,746	58.7	58.6
Low <= 15.8%	48,113	42,228	3,014	51.5	51.4

^a The information used to define the subgroups is from the 5-year (2009-2013) American Community Survey.

^b Finalized nonresponse includes refused cases, uncontacted cases known to be eligible, and other eligible cases not responding for reasons other than refusal.

^c Product of cases of unknown eligibility and estimated proportion of cases of unknown eligibility that are eligible.

^d Response rate = A/(A+B+C).

2.1.2 Extended Interviews

The PATH Study Adult Extended Interview and Youth Extended Interview collected extensive self-report information from the adult and youth sampled persons, respectively. The Adult Extended Interview gathered information from adults (18 years old and older) about tobacco use behaviors, attitudes, knowledge, and health conditions, as well as information on demographics, environmental factors, family and peer influences, substance use, and general physical and mental health status. Field interviewers conducted the Adult Extended Interviews in person using audio computer-assisted self-interviewing (ACASI). In Wave 1, 59,088 Phase 2 Screeners were finalized, and 32,400 Adult Extended Interviews were finalized.

The Youth Interview gathered information from youth (12 to 17 years old) on topics similar to those in the Adult Extended Interview. Sampled youth were asked about their tobacco use and attitudes about tobacco; in addition, youth were asked for information on demographics,

environmental factors, family and peer influences, substance use, and mental health. Field interviewers conducted the interviews in person using ACASI. The Study finalized 17,451 Youth Extended Interviews in Wave 1.

Method

Wave 1 response rates depended on completion of the Phase 1 Household Screener. For the Adult Extended Interview, the response rates were calculated as the product of (1) the Phase 2 or Individual Screener⁵ response rate; and (2) the proportion of adults who completed the Adult Extended Interview among those who completed the Phase 2 Screener and were selected for the Adult Extended Interview:

$$RRU = (C_{P2} / (C_{P2} + NR_{P2})) * (C_E / (C_E + NR_E))$$

where

- C_{P2} = number of completed cases or sufficient partials for the Phase 2 Screener;
- NR_{P2} = number of finalized nonresponse cases for the Phase 2 Screener;
- C_E = number of completed cases or sufficient partials for the Extended Interview; and
- NR_E = number of finalized nonresponse cases for the Extended Interview.

All cases were eligible.

For the Youth Extended Interview, the unweighted unit response rate is as follows:

$$RRU = C / (C + NR),$$

where

- C = number of completed cases or sufficient partials for the Extended Interview; and
- NR = number of finalized nonresponse cases for the Extended Interview.

Again, all cases were eligible.

⁵ Adults selected on the basis of the Phase 1 Household Screener were asked to complete the Phase 2 or Individual Screener. Those who completed the Phase 2 Screener were eligible for selection for the Adult Extended Interview, subject to further subsampling to achieve the design targets for the various age, race, and tobacco use groups. Of the adults who completed the Phase 2 Screener and were selected for the Adult Extended Interview, approximately 99.7 percent completed the Adult Extended Interview.

The adult and youth response rates were weighted to compensate for unequal probabilities of selection due to planned oversampling of individuals with certain characteristics. Person-level weights were computed as the product of the Household Screener IPS weights and individual IPS weights, which were calculated as the inverse of the selection probabilities for all persons sampled (responding and nonresponding persons). (See Section 2.2 for additional information on weighting.)

Tables 2-2 and 2-3 provide response rates for the Adult Extended Interview and Youth Extended Interview, respectively. In addition to the overall row, these tables include rows on age, sex, race, and ethnicity subgroups; Table 2-2 also includes a row on tobacco use status.⁶ Information from the Household Screener is used to define the demographic characteristics for the responding and nonresponding persons. Among adults for whom the Household Screener had missing values for information on tobacco use status, sampling at Phase 1 was based on the selection probabilities used for tobacco users, shown in the “sampled as user” row of Table 2-2. Adults for whom the Household Screener had missing values for other characteristics were excluded from the response rate calculation for the particular characteristic.

Results

For the **Adult Extended Interview**, the weighted overall response rate was 74.0 percent. (See Table 2-2.) This overall rate was lower than the projected rate of 85 percent, but it exceeded the worst-case scenario response rate of 58.1 percent previously provided to OMB in Attachment 22.

The weighted response rates for tobacco use status and demographic subgroups indicate the subgroups differed from one another by as much as 10.5 percentage points. As noted, information on the tobacco use status and demographic characteristics of eligible participants used in this table was gathered in the Household Screener. The differences among subgroups on weighted response rates were 3.7 percentage points for tobacco use status, 3.7 percentage points for age, 5.1 percentage points for sex, 10.5 percentage point for race, and 2.3 percentage points for ethnicity.

For the **Youth Extended Interview**, the weighted overall response rate was 78.4 percent. (See Table 2-3.) This overall rate was higher than the projected rate of 75 percent in Attachment 22. A worst-case scenario response rate was not specified for the Youth Interview.

⁶ Tobacco use status is based on information obtained in the Household Screener.

Table 2-2. PATH Study Wave 1 response rates by respondent characteristics: Adult Extended Interview

Characteristic ^a	Phase 2 Screener		Adult Extended Interview		Unweighted response rate for Wave 1 ^c (%)	Weighted response rate for Wave 1 (%)
	A: P2 Screener, completed (n)	B: P2 Screener, finalized nonresponse ^b (n)	C: Adult Extended, completed (n)	D: Adult Extended, finalized nonresponse ^b (n)		
Overall	44,303	14,785	32,320	80	74.8	74.0
Tobacco use status						
Sampled as user	23,859	7,299	21,362	55	76.4	76.4
Sampled as non-user	20,444	7,486	10,958	25	73.0	72.7
Age ^d						
18-24	11,679	3,732	9,059	16	75.6	75.1
25-44	14,663	4,593	11,266	18	76.0	75.4
45-64	12,503	4,229	8,761	30	74.5	73.8
65+	5,207	2,002	3,088	16	71.9	71.7
Sex ^d						
Male	21,546	8,166	16,321	50	72.3	71.4
Female	22,743	6,574	15,996	30	77.4	76.5
Race ^d						
White alone	32,222	11,085	23,645	53	74.2	73.8
Black alone or in combination	7,602	1,884	5,532	15	79.9	79.2
Other	3,344	1,363	2,452	8	70.8	68.7
Ethnicity ^d						
Hispanic	7,818	2,270	5,400	18	77.2	76.0
Non-Hispanic	36,430	12,433	26,891	62	74.4	73.7

^a The characteristics are as sampled. That is, information on the characteristics was collected in the Household Screener.

^b Finalized nonresponse includes refused cases and other eligible cases not responding for reasons other than refusal.

^c Response rate = $(A/(A+B)) * (C/(C+D))$.

^d The sum of counts for this category for Adult Extended, Completed (Column C) do not sum to the overall total due to missing values. The number of missing cases is 146 for age, 3 for sex, 691 for race, and 29 for ethnicity.

Table 2-3. PATH Study Wave 1 response rates by respondent characteristics: Youth Interview

Characteristic ^a	A: Completed (n)	B: Finalized nonresponse ^b (n)	Unweighted response rate for Wave 1 ^c (%)	Weighted response rate for Wave 1 (%)
Overall	13,651	3,800	78.2	78.4
Age ^d				
12-14	6,925	1,775	79.6	79.8
15-17	6,583	1,974	76.9	77.2
Sex ^d				
Male	6,988	1,960	78.1	78.2
Female	6,659	1,838	78.4	78.6
Race ^d				
White alone	9,623	2,705	78.1	78.2
Black alone or in combination	2,511	573	81.4	81.6
Other	1,079	408	72.6	72.7
Ethnicity ^d				
Hispanic	3,757	848	81.6	82.0
Non-Hispanic	9,873	2,947	77.0	77.1

^a The characteristics are as sampled. That is, information on the characteristics was collected in the Household Screener.

^b Finalized nonresponse includes refused cases and other eligible cases not responding for reasons other than refusal.

^c Response rate = A/(A+B).

^d The sum of counts for this category for Completed (Column A) do not sum to the overall total due to missing values. The number of missing cases is 143 for age, 4 for sex, 438 for race, and 21 for ethnicity.

The findings on the weighted response rates for demographic subgroups indicate the subgroups differed from one another by as much as 8.9 percentage points. Information on the demographic characteristics of eligible participants used in this table was gathered in the Household Screener. Differences among subgroups on weighted response rates were 2.6 percentage points for age, 0.4 percentage points for sex, 8.9 percentage point for race, and 4.9 percentage points for ethnicity.

2.1.3 Biospecimen Collections

This section is on the method and response rates for the collection of urine and blood samples from adults who completed Adult Extended Interviews.⁷ Biospecimens are intended to provide a basis for

⁷ Buccal cells were collected from adult participants for approximately 8 of the 15 months of Wave 1 (through May 18, 2014, when OMB approved a change request to discontinue the collection of this biospecimen). The PATH Study did not resume buccal cell collection in Wave 2; in addition, the Study has no plans to resume this collection in Wave 3.

the assessment of between-person differences and within-person changes in markers of tobacco exposure, and to detect and compare indicators of conditions and related disease processes associated with the use of tobacco products. Field interviewers collected the urine samples; on separate visits, phlebotomists collected the blood samples. The 32,320 adults who completed the Adult Extended Interview were eligible to provide biospecimens.

Method

Table 2-4 provides overall unweighted and weighted response rates for the biospecimen collections, and response rates for tobacco use status and demographic subgroups. The response rates are conditional on a completed Household Screener and a completed Adult Extended Interview. The response rates were calculated using the following formula:

$$\text{RRU} = \text{Number of samples collected/number of Adult Extended Interviews completed}$$

This is the same formula used to compute the projected biospecimen response rates presented in Attachment 22 for Wave 1. The denominator for the rate, the 32,320 adults who completed the Adult Extended Interview, is the same for both of the biospecimen response rates.

In addition to the overall row, the table includes rows on tobacco use status, age, sex, race, and ethnicity subgroups. Information from the Adult Extended Interview is used to define the tobacco use status and demographic characteristics for the responding and nonresponding adults. Adults with missing values for such characteristics were excluded from the response rate calculation for that characteristic.

Results

The weighted response rate for **urine** was 63.6 percent; the projected response rate was 80 percent, and the worst-case response rate was 49 percent. The differential weighted response rate for subgroups of respondents ranges from 2.2 percentage points for sex to 11.5 percentage points for age. The response rate for urine collection was lower than projected, but it exceeds the worst-case scenario discussed in Attachment 22.

Table 2-4. PATH Study Wave 1 response rates by respondent characteristics: Biospecimen collections

Characteristic ^a	A: Adult Extended Interviews completed (n)	Biospecimen					
		Urine			Blood		
		B: Collected (n)	Unweighted response rate for Wave 1 ^c	Weighted response rate for Wave 1	B: Collected (n)	Unweighted response rate for Wave 1 ^c	Weighted response rate for Wave 1
Overall	32,320	21,801	67.5	63.6	14,520	44.9	43.0
Tobacco use status ^b							
Sampled as user	23,084	16,116	69.8	68.9	10,764	46.6	46.4
Sampled as non-user	9,234	5,683	61.5	60.3	3,754	40.7	40.8
Age ^b							
18-24	9,110	6,457	70.9	69.0	3,884	42.6	41.4
25-44	11,269	7,742	68.7	65.8	5,004	44.4	41.0
45-64	8,818	5,725	64.9	62.2	4,191	47.5	45.4
65+	3,110	1,873	60.2	57.5	1,438	46.2	43.9
Sex ^b							
Male	16,309	10,763	66.0	62.4	6,920	42.4	41.1
Female	15,982	11,025	69.0	64.6	7,594	47.5	44.6
Race ^b							
White alone	23,242	15,531	66.8	63.2	10,637	45.8	44.2
Black alone or in combination	5,538	3,911	70.6	66.6	2,381	43.0	40.3
Other	2,726	1,801	66.1	61.4	1,152	42.3	38.4
Ethnicity ^b							
Hispanic	5,536	3,870	69.9	66.4	2,452	44.3	40.5
Non-Hispanic	26,288	17,633	67.1	63.1	11,885	45.2	43.7

^a The characteristics are as reported in the Adult Extended Interview.

^b The sum of counts for this category for Urine (Column B) do not sum to the overall total due to missing values. The number of missing cases is 2 for tobacco use, 4 for age, 13 for sex, 558 for race, and 298 for ethnicity. Also, the sum of counts for this category for Blood (Column B) do not sum to the overall total due to missing values. The number of missing cases is 2 for tobacco use, 3 for age, 6 for sex, 350 for race, and 183 for ethnicity.

^c Response rate = B/A.

The weighted response rate for **blood** was 43.0 percent; the projected response rate was 65 percent, and the worst-case response rate was 39 percent. The differential weighted response rate for subgroups of respondents ranges from 3.2 percentage points for ethnicity to 5.8 percentage points for race. The response rate for blood collection was lower than projected, but it exceeds the worst-case scenario in Attachment 22.

2.2 Nonresponse Bias Analysis

This nonresponse bias analysis investigates possible differences between estimates calculated from Wave 1 of the PATH Study and independent estimates of those quantities from other surveys and censuses. By so doing, the Study can assess the extent to which differential nonresponse among population subgroups may affect estimates. Results are presented on the characteristics of respondents to the Household Screener, Adult Extended Interview, and Youth Interview, and on adults from whom biospecimens were collected for the PATH Study. The analysis of nonresponse bias is based on the full set of Wave 1 data.

2.2.1 Method

The method used in the PATH Study to assess potential nonresponse bias begins by comparing estimates of demographic counts from the Wave 1 sample⁸ with corresponding estimates from the American Community Survey (ACS). The 1-year (2013) ACS estimates, calculated from the 2013 ACS Public Use Microdata Sample (PUMS),⁹ were used for comparison purposes. The estimates calculated from the ACS PUMS excluded institutional group quarters and persons in noninstitutional group quarters who were not college students. These exclusions correspond to exclusions from the target population for the PATH Study.

⁸ The full Wave 1 sample was used for the analyses in this section; however, the weights used for this report are preliminary. Estimates based on final releases of the Wave 1 PATH Study data may differ slightly from those in this report.

⁹ The ACS PUMS files provide a probability sample of individual records from the full ACS (United States Census Bureau, 2015), allowing data users to create custom tables that are not available through pre-tabulated ACS data products. Using the PUMS files allowed comparison estimates to be calculated that corresponded to the target population for the PATH Study.

The PATH Study measures a range of tobacco use behaviors; many of these variables are not available in other studies. However, responses to the PATH Study questions on current cigarette smoking can be compared with estimates from other surveys that ask about cigarette smoking behavior. The following surveys were used for comparison: the Tobacco Use Supplement to the Current Population Survey, 2010-2011 (TUS-CPS); the National Health and Nutrition Examination Survey, 2011-2012 (NHANES); the National Health Interview Survey, 2013 (NHIS); the National Survey on Drug Use and Health, 2013 (NSDUH); and the National Youth Tobacco Survey, 2012 (NYTS). Appendix A describes the questions used to define current smoking on each of these surveys as well as the PATH Study, and outlines differences in target populations among these surveys and the PATH Study.

The PATH Study oversamples young adults, African-American adults, and adult tobacco users. Consequently, unweighted estimates of population quantities would be expected to be biased. In this section, the inverse-probability-of-selection (IPS) weights, calculated using the probabilities of selection, are used to estimate population quantities. Without nonresponse, estimates calculated using the IPS weights would be expected to accord with the population counts.

The IPS weights were calculated in two stages. First, the household-level IPS weights (HHIPSWT) were calculated for all households sampled (responding households and nonresponding households) as the inverse of the selection probability:

$$HHIPSWT_{ijk} = \frac{1}{P_{ijk}}.$$

Here, P_{ijk} is the probability that household k in segment j of PSU i is selected to be in the sample. Addresses were sampled directly from the USPS CDSF, so that P_{ijk} is the product of the PSU, the segment-within-PSU, and the address-within-segment selection probabilities.

For nonresponse bias assessment purposes, the person-level IPS weights were computed using HHIPSWT. For youth ages 12-17, these were calculated as

$$YIPSWT_{ijkl} = HHIPSWT_{ijk} \times \frac{1}{\text{Probability youth } l \text{ in household } (ijk) \text{ selected for sample}}.$$

Most selected households had fewer than 3 youths, who were then selected with certainty, so that for most households, the youth IPS weight (YIPSWT) is the same as the household-level IPS weight.

Adults were selected with different probabilities according to their age, race, and tobacco use status. The adult IPS weights (AIPSWT) were calculated as

$$AIPSWT_{ijkl} = HHIPSWT_{ijk} \times \frac{1}{\text{Probability adult } l \text{ in household } (ijk) \text{ selected for sample}}$$

Sampling of adults occurred in two phases. Phase 1 selected adults based on responses to the Household Screener. The probability that adult l in the household was selected in Phase 1 was a function of the number of adults in the household and of the ages, races, and tobacco use statuses reported for those adults by the informant completing the Household Screener. Adults sampled at Phase 1 were individually asked questions about their age, race, and tobacco use as part of the Phase 2 Screener, and were subsampled for continuation into the Adult Extended Interview on the basis of their responses to those questions. The probability in the formula for AIPSWT is the product of the first-phase and second-phase selection probabilities.

Note that no nonresponse adjustments were performed for the calculation of IPS weights. The weights HHIPSWT, YIPSWT, and AIPSWT were used for all calculations employing IPS weights reported in Section 2.2.2. In the tables presented in Section 2.2.2, the unweighted counts include categories for missing values. Estimates of percentages calculated using weights, however, exclude respondents with missing values for that item. The estimates calculated from other surveys that are used for comparison purposes also exclude missing values, except where noted.

Point estimates for the PATH Study were calculated using HHIPSWT for household estimates, AIPSWT for adult estimates, and YIPSWT for youth estimates. Replicate weights, used to calculate variances, account for the complex sampling features of stratification and clustering. Precisions for the estimates are reported using 95 percent confidence intervals. Estimates from TUS-CPS, NHANES, NHIS, NSDUH, and NYTS¹⁰ also have sampling error, so 95 percent confidence

¹⁰ Estimates from TUS-CPS were obtained from US Department of Commerce (2012) published tables. Estimates from NHANES, NHIS, and NYTS were calculated from their respective public use files (CDC, 2014a, 2014b, 2015b). Estimates from NSDUH were obtained from CBHSQ (2014), SAMHSA (2014), and from the Survey Documentation and Analysis system at <http://www.icpsr.umich.edu/cgi-bin/SDA/SAMHDA/hsda?samhda+35509-0001>.

intervals are reported for the estimates from those surveys as well.¹¹ SAS software version 9.3 (SAS Institute, 2011) was used to calculate all point estimates and confidence intervals.

2.2.2 Results

The first set of tables provides estimates derived from the Household Screener. Demographic characteristics were estimated using the information on the roster of household members provided by the household respondent. The household-level IPS weight HHIPSWT was used in Tables 2-5 through 2-9 to evaluate potential nonresponse bias. If nonresponse is not associated with demographic characteristics, then the percentages calculated using HHIPSWT would be expected to be close to those from the ACS. Note that the ACS estimates are for calendar year 2013 while the PATH Study Wave 1 data were collected between September, 2013 and December, 2014. There is therefore the possibility that differences for some rapidly changing characteristics may reflect the later data collection for the PATH Study.

Table 2-5 presents unweighted counts and estimated population percentages for adults in the four race/age domains used for sampling adults within households based on the enumeration of adults from the Household Screener. Comparisons of the PATH Study estimates to the ACS estimates for these four domains are shown in Table 2-5. The IPS-weighted estimates of percentages in the Black 18-24 and Black 25+ domains are similar to the 1-year 2013 ACS estimates. The PATH Study estimate of the percentage in the non-Black 18-24 domain is 1.3 percentage points higher than the corresponding ACS estimate, and the PATH Study estimate in the non-Black 25+ domain is 2.6 percentage points lower than the ACS estimate.

¹¹ If the confidence interval from the PATH Study estimate does not overlap with the confidence interval from the comparison study, then the results are significantly different at the 0.05 significance level. Schenker and Gentleman (2001) show that this results in a conservative test. Confidence intervals are not reported for the ACS results because they are based on such large sample sizes that the sampling error is negligible. For comparison with ACS results, the PATH Study estimate significantly differs from the ACS estimate at the 0.05 significance level if the ACS percentage is outside of the PATH Study confidence interval. In general, if a 95 percent confidence interval for percentage of adults who are current smokers from the PATH Study includes a fixed value X, then a hypothesis test of the null hypothesis that the percentage of adults who are current smokers equals X would have p-value > 0.05 and therefore the difference between the PATH Study estimate and the estimate from the external survey is not statistically significant.

Table 2-5. Race by age distribution, based on the household enumeration: IPS weights

Race and age classification	Unweighted count	Weighted percentage, using household IPS weights	Confidence interval using household IPS weights	Percentage from ACS PUMS
Black* 18-24	3,669	2.3%	[2.0%, 2.6%]	2.0%
Black* 25+	17,934	11.4%	[9.8%, 12.9%]	10.5%
Non-Black 18-24	18,058	11.4%	[10.8%, 12.0%]	10.1%
Non-Black 25+	116,957	74.9%	[73.3%, 76.5%]	77.5%
Missing age or race	4,003			
Total	160,621	100.0%		100.0%

*Black alone or in combination with other race(s).

Table 2-6 compares the sex of the adults enumerated on the PATH Study household rosters with the 1-year 2013 ACS distribution. The confidence intervals from the PATH Study include the ACS percentages, indicating no evidence of nonresponse bias with respect to sex.

Table 2-6. Distribution of male and female adults listed in the household enumeration: IPS weights

Sex	Unweighted count	Weighted percentage for adults, using household IPS weights	Confidence interval using household IPS weights	Percentage from ACS PUMS
Male	77,088	48.0%	[47.8%, 48.3%]	48.2%
Female	83,500	52.0%	[51.7%, 52.2%]	51.8%
Missing	33			
Total	160,621	100.0%		100.0%

Table 2-7 compares the distribution of household size for the responding households with corresponding estimates from the 1-year 2013 ACS. Compared with the ACS, the PATH Study estimates fewer single- and two-person households and more households with at least three persons. The PATH Study also estimates a lower percentage of single-adult households (Table 2-8) and, probably related to this pattern, a slightly higher estimated percentage of households with youth ages 12-17 than found in the ACS (Table 2-9).¹² If no further weighting adjustments were performed then to the extent that household size is associated with the PATH Study’s outcomes, those outcomes may be affected by nonresponse bias. However, this concern is addressed by the weighting adjustments and results described in Section 2.3.

¹² Surveys commonly achieve a slightly lower percentage of one-person households because they have fewer members available for contact. See Brault (2013), who found a similar pattern in the CPS ASEC content test.

Table 2-7. Distribution of household size based on households responding to the Household Screener: IPS weights

Number of persons in household	Unweighted count	Weighted percentage, using household IPS weights	Confidence Interval using household IPS weights	Percentage from ACS PUMS
1	17,822	22.7%	[22.1%, 23.4%]	27.7%
2	26,098	33.0%	[32.5%, 33.5%]	33.7%
3	13,503	17.0%	[16.7%, 17.3%]	15.7%
4	11,720	14.7%	[14.3%, 15.1%]	13.0%
5+	10,055	12.6%	[12.0%, 13.2%]	9.9%
Total	79,198	100.0%		100.0%

Table 2-8. Distribution of number of adults based on households responding to the Household Screener: IPS weights

Number of adults in household	Unweighted count	Weighted percentage, using household IPS weights	Confidence Interval using household IPS weights	Percentage from ACS PUMS
0-1*	21,869	27.8%	[27.1%, 28.5%]	33.3%
2	40,961	51.7%	[51.2%, 52.2%]	50.7%
3+	16,368	20.5%	[19.8%, 21.1%]	16.0%
Total	79,198	100.0%		100.0%

*A small number of households contain only emancipated youth and hence contribute to the zero part of this category.

Table 2-9. Distribution of number of youth ages 12-17 based on households responding to the Household Screener: IPS weights

Number of youth ages 12-17 in household	Unweighted count	Weighted percentage, using household IPS weights	Confidence interval using household IPS weights	Percentage from ACS PUMS
0	65,661	83.0%	[82.4%, 83.5%]	84.4%
1	9,569	12.0%	[11.7%, 12.4%]	11.1%
2+	3,968	5.0%	[4.7%, 5.2%]	4.5%
Total	79,198	100.0%		100.0%

Tables 2-10 and 2-11 are based on adults in the Wave 1 sample responding to the Adult Extended Interview. As noted above, the PATH Study oversamples young adults, African-American adults, and adult tobacco users, so estimates calculated without weights will not accord with population estimates. The IPS-weighted estimates are calculated using the adult weight AIPSWT; if the PATH Study had full response, it would be expected that the IPS-weighted estimates would be close to the corresponding population quantities. Table 2-10 presents the estimated race, ethnicity, and sex/age distributions from adults in the Wave 1 sample responding to the Adult Extended Interview.

Table 2-10. Demographic distributions based on adults responding to the Adult Extended Interview, and based on adults providing urine or blood specimens: IPS weights

	Adult respondents to Adult Extended Interview			Adults from whom urine specimen is collected			Adults from whom blood specimen is collected			ACS PUMS
	Un-weighted count	Weighted percentage, using adult IPS weights	Confidence interval using adult IPS weights	Un-weighted count	Weighted percentage, using adult IPS weights	Confidence interval using adult IPS weights	Un-weighted count	Weighted percentage, using adult IPS weights	Confidence interval using adult IPS weights	Weighted percentage
Sex										
Male	16,309	45.7%	[44.9%, 46.6%]	10,763	44.9%	[43.9%, 45.9%]	6,920	43.7%	[42.6%, 44.8%]	48.1%
Female	15,982	54.3%	[53.4%, 55.1%]	11,025	55.1%	[54.1%, 56.1%]	7,594	56.3%	[55.2%, 57.4%]	51.9%
Missing	29			13			6			
Total	32,320	100.0%		21,801	100.0%		14,520	100.0%		100.0%
Age group										
18-24	9,110	13.8%	[13.2%, 14.4%]	6,457	14.9%	[14.2%, 15.6%]	3,884	13.3%	[12.5%, 14.0%]	13.0%
25-44	11,280	35.8%	[35.0%, 36.6%]	7,744	37.0%	[36.0%, 38.0%]	5,005	34.1%	[33.1%, 35.2%]	34.3%
45-64	8,818	33.7%	[32.8%, 34.5%]	5,725	33.0%	[32.0%, 33.9%]	4,191	35.6%	[34.4%, 36.7%]	34.5%
65+	3,110	16.7%	[16.0%, 17.4%]	1,873	15.1%	[14.3%, 15.9%]	1,438	17.1%	[16.1%, 18.1%]	18.2%
Missing	2			2			2			
Total	32,320	100.0%		21,801	100.0%		14,520	100.0%		100.0%
Race										
Black alone or in combination	5,538	14.1%	[12.1%, 16.1%]	3,911	14.8%	[12.6%, 17.0%]	2,381	13.2%	[11.2%, 15.2%]	12.5%
White alone	23,242	77.4%	[75.1%, 79.7%]	15,531	77.0%	[74.6%, 79.4%]	10,637	79.3%	[77.1%, 81.5%]	75.7%
Other	2,726	8.5%	[7.5%, 9.5%]	1,801	8.2%	[7.2%, 9.2%]	1,152	7.5%	[6.7%, 8.4%]	11.8%
Missing	814			558			350			
Total	32,320	100.0%		21,801	100.0%		14,520	100.0%		100.0%
Ethnicity										
Hispanic	5,536	17.8%	[15.3%, 20.3%]	3,870	18.6%	[16.0%, 21.2%]	2,452	16.7%	[13.9%, 19.5%]	15.0%
Non-Hispanic	26,288	82.2%	[79.7%, 84.7%]	17,633	81.4%	[78.8%, 84.0%]	11,885	83.3%	[80.5%, 86.1%]	85.0%
Missing	496			298			183			
Total	32,320	100.0%		21,801	100.0%		14,520	100.0%		100.0%

Table 2-11. Comparison of education level and health insurance status based on adults completing the Adult Extended Interview and based on adults providing urine or blood specimens: IPS weights

	Adult respondents to Adult Extended Interview			Adults from whom urine specimen collected			Adults from whom blood specimen collected			ACS PUMS
	Un-weighted count	Weighted percentage, using adult IPS weights	Confidence interval using adult IPS weights	Un-weighted count	Weighted percentage, using adult IPS weights	Confidence interval using adult IPS weights	Un-weighted count	Weighted percentage, using adult IPS weights	Confidence interval using adult IPS weights	Weighted percentage
Education										
18-24										
< HS, HS or GED	4,465	49.9%	[48.1%, 51.7%]	3,234	50.4%	[48.3%, 52.5%]	1,948	50.2%	[47.7%, 52.6%]	43.6%
> HS	4,593	50.1%	[48.3%, 51.9%]	3,197	49.6%	[47.5%, 51.7%]	1,928	49.8%	[47.4%, 52.3%]	56.4%
Subtotal	9,058	100.0%		6,431	100.0%		3,876	100.0%		100.0%
25+										
< HS or GED	4,769	18.0%	[16.9%, 19.1%]	3,425	19.3%	[18.1%, 20.6%]	2,405	19.0%	[17.6%, 20.5%]	16.9%
HS	4,763	20.0%	[19.1%, 21.0%]	3,106	19.8%	[18.6%, 20.9%]	2,071	19.0%	[17.7%, 20.2%]	23.8%
Some college, no degree	7,650	31.3%	[30.2%, 32.3%]	5,238	32.4%	[31.3%, 33.4%]	3,721	33.4%	[31.9%, 34.8%]	29.2%
Bachelor degree	3,642	18.2%	[17.1%, 19.2%]	2,139	16.7%	[15.6%, 17.9%]	1,454	16.7%	[15.3%, 18.1%]	18.7%
> Bachelor degree	2,228	12.5%	[11.5%, 13.5%]	1,369	11.8%	[10.8%, 12.8%]	956	11.9%	[10.8%, 13.1%]	11.3%
Subtotal	23,052	100.0%		15,277	100.0%		10,607	100.0%		100.0%
Missing	210			93			37			
Total	32,320			21,799			14,520			
Health insurance										
Yes	25,760	84.4%	[83.5%, 85.2%]	17,306	83.9%	[83.0%, 84.8%]	11,686	85.0%	[84.0%, 86.0%]	83.3%
No	6,156	15.6%	[14.8%, 16.5%]	4,335	16.1%	[15.2%, 17.0%]	2,758	15.0%	[14.0%, 16.0%]	16.7%
Missing	404			160			76			
Total	32,320	100.0%		21,799	100.0%		14,520	100.0%		100.0%

Additional columns in the table present the weighted distributions, using weight AIPSWT, for the adults from whom urine and/or blood specimens were collected.

When using unweighted counts, males comprise more than 50 percent of the 32,320 PATH Study respondents in Wave 1. By contrast, the IPS-weighted estimated percent of males in the population at Wave 1 is significantly lower than the ACS estimate. This is also true for the subset of respondents to the Adult Extended Interview who provided blood and/or urine specimens. Age group comparisons show that the IPS-weighted estimates of the percent of adults who are ages 18-44 based on the PATH Study respondents to the Adult Extended Interview, and based on the adults providing urine specimens, are significantly higher than the ACS estimate. The nonresponse-adjusted weights in Section 2.3, which calibrate to age groups and to sex, correct for these discrepancies for the Adult Extended Interview respondents.

Table 2-10 shows that the estimated percentages in different race and ethnicity groups, calculated using adults responding to the Adult Extended Interview, or using those who provide blood specimens, are similar to the 1-year 2013 ACS estimates of those quantities, with the exceptions that persons in the “other race” category are underrepresented among respondents to the Adult Extended Interview and persons providing urine and/or blood specimens, with corresponding slight overrepresentation of Whites among the persons providing blood specimens and of Blacks among the persons providing urine specimens. Hispanics are slightly overrepresented among respondents to the Adult Extended Interview and persons providing urine specimens.

Table 2-11 compares Adult Extended Interview respondents and those from whom biospecimens were collected to the ACS with respect to education level and presence of health insurance. The estimated distributions of educational level from adults responding to the Adult Extended Interview and from adults providing biospecimens differ from that in the ACS. For adults responding to the interview and for adults providing biospecimens, the estimated percentages of 18-24 year-olds with at least some college are lower than the ACS; among older adults, the estimated percentages of persons with a high school degree are lower than the ACS and the estimated percentages of persons with some college but no degree are higher than the ACS. In addition, among the adults providing biospecimens the estimated percentages of adults ages 25+ with less than a high school education are higher than the ACS and the estimated percentages of adults ages 25+ with a bachelor’s degree are lower than the ACS. Education level has been shown to be associated with tobacco use status (Agaku et al., 2014); the nonresponse-adjusted weights described in Section 2.3 adjust for educational attainment. Estimates of percentage of adults with health insurance based on respondents to the Adult Extended Interview and on those who provided blood samples are higher

than the corresponding estimate from the 2013 ACS, although no significant difference was found between the estimate from the PATH Study respondents providing urine samples and the estimate from the ACS. It should be noted, however, that the ACS estimates do not account for persons who obtained insurance in 2014 under the Affordable Care Act.

Table 2-12 presents the estimates of prevalence of current cigarette smoking¹³ for adults based on the Adult Extended Interview, for the adult population as a whole and for subgroups. These estimates are accompanied by 95 percent confidence intervals for the percentage of current cigarette smokers for the PATH Study estimates. The last five columns are the estimates of smoking prevalence from TUS-CPS, NHIS, NHANES, and NSDUH, respectively, along with 95 percent confidence intervals from those surveys. Note that these estimates exclude responses of “don’t know” and missing values.

The estimates of current smoking prevalence differ substantially from survey to survey. Many potential reasons can explain these disparities, including that each survey has sampling error. Beyond that, however, the surveys differ in question order, context, design, mode of administration, and year of most recent data collection.

In general, the TUS-CPS estimates of smoking prevalence are lower than estimates from the other surveys, including the PATH Study. This may be related to the proxy responses used in the TUS-CPS. The rotation group structure of the TUS-CPS may result in underestimates of smoking prevalence, as smokers are more likely to drop out over the course of the panel survey (Song, 2013).

The PATH Study and NSDUH both use ACASI administration for the tobacco use questions so that the interviewer does not see responses to the questions. By contrast, TUS-CPS, NHIS, and NHANES have direct questioning by an interviewer: NHIS and NHANES are conducted in person, and TUS-CPS is conducted in person and by telephone. The contexts and purposes of these surveys also differ: CPS is a general survey on unemployment, NHIS and NHANES are general health surveys, and NSDUH is a cross-sectional survey on substance use (including tobacco use) and health, including mental health. Unlike the cross-sectional prevalence surveys, the PATH Study is designed for research purposes and uses a longitudinal cohort design to assess within-person change

¹³ For the PATH Study, following common practice for tobacco surveys, a current smoker is someone who (1) has smoked at least 100 cigarettes in his or her lifetime and (2) currently smokes every day or some days. The questions used to define current smoking for each survey are provided in Appendix A.

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.

Table 2-12 indicates the IPS-weighted estimates of current smoking from the PATH Study are most similar to estimates from NHIS and NHANES. All of the 95 percent confidence intervals for percent of current cigarette smokers constructed from the PATH Study overlap with the confidence intervals for NHIS, NHANES, or both. Estimates from TUS-CPS tend to be below the estimates from the PATH Study, NHIS, and NHANES; estimates from NSDUH tend to be above the estimates from the PATH Study, NHIS, and NHANES. No evidence was found to indicate nonresponse bias in the PATH Study with respect to cigarette smoking behavior among adults, because the PATH Study's estimates are all within the range of estimates from comparable surveys.

Table 2-13 gives estimates of current cigarette smoking for the adults from whom urine and/or blood specimens were collected. The IPS-weighted estimates of smoking are higher for adults who contributed biospecimens. However, the confidence intervals for smoking among adults providing biospecimens are in line with the estimates from NSDUH and NHANES.

Results in Tables 2-10 through 2-13 are based on adults in the full Wave 1 sample responding to the Adult Extended Interview. Similar analyses were performed for the youth respondents. The demographic estimates are given in Table 2-14 and estimates of two measures of cigarette smoking commonly applied to youth are given in Table 2-15.

Table 2-14 shows that the IPS-weighted estimates of percentages of youth who are male/female and ages 12-13/14-17 are not significantly different from the 1-year 2013 ACS percentages. The PATH Study estimate of the percent of youth who are Hispanic, however, is approximately 6 percentage points higher than the corresponding estimate from ACS, indicating that Hispanic youth were more likely to respond to the PATH Study survey.

Table 2-12. Current cigarette smoking based on adults responding to the Adult Extended Interview: IPS weights

	Sample size	PATH Study: Unweighted percentage	PATH Study: Weighted percentage, using adult IPS weights [95% confidence interval]	Percentage from 2010-2011 TUS-CPS [95% confidence interval]	Percentage from 2013 NHIS [95% confidence interval]	Percentage from 2011-2012 NHANES* [95% confidence interval]	Percentage from 2013 NSDUH, original definition** [95% confidence interval]	Percentage from 2013 NSDUH, modified definition [95% confidence interval]
Current smoker	32,245	35.4%	18.6% [17.7%, 19.4%]	16.1% [15.8%, 16.3%]	17.8% [17.2%, 18.4%]	19.8% [17.5%, 22.1%]	22.9% [22.6%, 23.2%]	21.0% [20.4%, 21.7%]
Current smoker, male	16,265	36.8%	21.2% [20.3%, 22.2%]	18.0% [17.7%, 18.4%]	20.5% [19.5%, 21.4%]	23.9% [20.7%, 27.1%]	25.6% [25.1%, 26.1%]	23.8% [22.7%, 24.9%]
Current smoker, female	15,952	33.9%	16.4% [15.4%, 17.3%]	14.2% [13.9%, 14.5%]	15.3% [14.6%, 16.0%]	16.0% [13.5%, 18.5%]	20.4% [20.0%, 20.8%]	18.4% [17.6%, 19.3%]
Current smoker, age 18-24	9,099	27.2%	20.1% [18.8%, 21.5%]	17.1% [16.4%, 17.8%]	18.7% [16.8%, 20.5%]	20.4%*** [13.7%, 27.1%]	NA****	NA
Current smoker, age 25-44	11,260	41.2%	22.9% [21.8%, 24.1%]	17.9% [17.5%, 18.4%]	20.1% [19.1%, 21.1%]	23.3% [20.0%, 26.7%]	NA	NA
Current smoker, age 45-64	8,784	40.6%	18.7% [17.7%, 19.8%]	17.8% [17.4%, 18.2%]	20.0% [19.0%, 20.8%]	21.3% [18.3%, 24.2%]	NA	NA
Current smoker, age 65+	3,100	23.3%	7.7% [7.0%, 8.4%]	7.8% [7.5%, 8.2%]	8.8% [8.0%, 9.6%]	9.2% [6.7%, 11.7%]	NA	NA
Current smoker, Hispanic	5,519	26.3%	13.4% [12.6%, 14.2%]	10.9% [10.4%, 11.5%]	12.1% [11.0%, 13.2%]	16.6% [13.7%, 19.5%]	18.9% [18.1%, 19.7%]	15.2% [13.8%, 16.7%]
Current smoker, white non-Hispanic	19,268	38.8%	19.6% [18.5%, 20.7%]	17.5% [17.2%, 17.8%]	19.4% [18.5%, 20.2%]	20.2% [17.0%, 23.3%]	24.1% [23.7%, 24.5%]	22.8% [21.9%, 23.6%]
Current smoker, other non-Hispanic	6,904	33.3%	20.4% [19.3%, 21.5%]	NA	16.7% [15.6%, 17.8%]	20.8% [16.6%, 24.9%]	21.8% [20.2%, 23.4%]	19.6% [18.1%, 21.2%]
Current every-day smoker	32,245	28.0%	14.8% [13.9%, 15.6%]	12.7% [12.4%, 12.9%]	13.7% [13.1%, 14.2%]	16.4% [14.3%, 18.4%]	NA	NA
Current some-days smoker	32,245	7.4%	3.8% [3.6%, 4.1%]	3.4% [3.3%, 3.5%]	4.1% [3.8%, 4.4%]	3.4% [2.7%, 4.1%]	NA	NA

*The smoking questions asked in NHANES for adults ages 20 and older differ from the questions asked for persons ages 12-19. The modes of administration also differ for the two age groups. The NHANES estimates presented in this table are for adults ages 20 and older.

**NSDUH's definition of a current cigarette smoker is someone who has smoked part or all of a cigarette in the past 30 days, which is more expansive than the definition used in the other surveys. However, NSDUH contains questions on lifetime smoking and current smoking. The modified definition uses these questions to construct a measure of "current smoking" that is comparable to that of the other surveys (Ryan et al., 2012). The construction of this variable is described in Appendix A. The estimates and confidence intervals for the NSDUH "original definition" (except for the "current smoker, other non-Hispanic" estimate) are from the published tables (SAMHSA, 2014); the estimates and confidence intervals for the "modified definition" are calculated from the public use data set. The estimate of current smoking for the "other non-Hispanic" group was not available from the published tables and it was also calculated from the public use data set.

***The estimate is for adults 20-24 years old.

****Detailed age information was not available in the public use file for NSDUH 2013.

Table 2-13. Current cigarette smoking based on adults providing biospecimens in Wave 1: IPS weights

	Sample size	PATH Study: Weighted cigarette smoking prevalence, using adult IPS weights [95% confidence interval]	Percentage from 2010-2011 TUS-CPS [95% confidence interval]	Percentage from 2013 NHIS [95% confidence interval]	Percentage from 2011-2012 NHANES [95% confidence interval]	Percentage from 2013 NSDUH, original definition* [95% confidence interval]	Percentage from 2013 NSDUH, modified definition [95% confidence interval]
Adult respondent to Adult Extended Interview	32,245	18.6% [17.7%, 19.4%]	16.1% [15.8%, 16.3%]	17.8% [17.2%, 18.4%]	19.8% [17.5%, 22.1%]	22.9% [22.6%, 23.2%]	21.0% [20.4%, 21.7%]
Adults providing urine	21,757	20.9% [20.0%, 21.9%]	16.1% [15.8%, 16.3%]	17.8% [17.2%, 18.4%]	19.8% [17.5%, 22.1%]	22.9% [22.6%, 23.2%]	21.0% [20.4%, 21.7%]
Adults providing blood	14,493	21.6% [20.5%, 22.7%]	16.1% [15.8%, 16.3%]	17.8% [17.2%, 18.4%]	19.8% [17.5%, 22.1%]	22.9% [22.6%, 23.2%]	21.0% [20.4%, 21.7%]

*NSDUH's definition of a current cigarette smoker is someone who has smoked part or all of a cigarette in the past 30 days. However, NSDUH contains questions on lifetime smoking and current smoking. The modified definition uses these questions to construct a measure of "current smoking" that is comparable to that of the other surveys (Ryan et al., 2012). The construction of this variable is described in Appendix A.

Table 2-14. Demographic distributions based on youth ages 12-17 who completed the Youth Interview: IPS weights

	Unweighted count	Weighted percentage, using youth IPS weights	Confidence interval using youth IPS weights	Percentage from ACS PUMS
Sex				
Male	6,971	51.2%	[50.3%, 52.1%]	51.3%
Female	6,641	48.8%	[47.9%, 49.7%]	48.7%
Missing	39			
Total	13,651	100.0%		100.0%
Age group				
12-13	4,684	34.3%	[33.5%, 35.2%]	33.7%
14-17	8,965	65.7%	[64.8%, 66.5%]	66.3%
Missing	2			
Total	13,651	100.0%		100.0%
Race/ethnicity				
Hispanic	3,880	28.6%	[25.2%, 32.1%]	22.3%
Non-Hispanic white alone	6,616	48.4%	[45.1%, 51.7%]	54.5%
Non-Hispanic other	3,135	23.0%	[20.5%, 25.5%]	23.2%
Missing	20			
Total	13,651	100.0%		100.0%

Table 2-15 provides estimates from the PATH Study for two common measures of cigarette smoking prevalence among youth respondents compared with estimates from NHANES, NSDUH, and NYTS.¹⁴ Different measures of smoking are used in this report for youth than for adults. The primary measure of cigarette smoking among youth in this report is whether the youth has ever tried smoking a cigarette, even one or two puffs (see Appendix A). Another measure is current smoking, defined as having smoked at all in the past 30 days. Both are shown in Table 2-15.

¹⁴ TUS-CPS does not interview persons younger than 18 about tobacco use.

Table 2-15. Cigarette smoking* based on youth ages 12-17 who completed the Youth Interview: IPS weights

	Sample size	PATH Study: Unweighted percentage	PATH Study: Weighted percentage, using youth IPS weights [95% confidence interval]	Percentage from 2011-2012 NHANES [95% confidence interval]	Percentage from 2013 NSDUH [95% confidence interval]	Percentage from 2012 NYTS [95% confidence interval]
Ever tried cigarette smoking, even one or two puffs	13,631	13.5%	13.5% [12.6%, 14.5%]	20.5% [17.5%, 23.6%]	15.7% [15.4%, 16.0%]	25.6% [23.6%, 27.6%]
Ever tried smoking, male	6,959	13.9%	14.0% [13.0%, 15.1%]	21.1% [15.9%, 26.3%]	16.3% [15.8%, 16.8%]	27.2% [25.0%, 29.3%]
Ever tried smoking, female	6,634	13.1%	13.1% [12.0%, 14.2%]	20.0% [14.6%, 25.5%]	15.1% [14.6%, 15.6%]	24.0% [21.8%, 26.2%]
Ever tried smoking, age 12-13	4,675	4.6%	4.6% [3.9%, 5.4%]	5.6% [1.9%, 9.4%]	4.0% [3.7%, 4.3%]	11.8% [10.2%, 13.4%]
Ever tried smoking, age 14-17	8,954	18.1%	18.2% [17.1%, 19.4%]	28.3% [23.5%, 33.0%]	21.0% [20.0%, 22.0%]	32.5% [30.0%, 34.9%]
Have smoked in past 30 days	13,613	4.7%	4.7% [4.2%, 5.1%]	6.9% [4.0%, 9.8%]	5.6% [5.4%, 5.8%]	8.7% [7.7%, 9.8%]

*Defined as ever tried a cigarette, even one or two puffs. For comparison, an additional measure of current smoking commonly applied to youth (having smoked at all in the past 30 days) is also included in this table.

Differences in target populations and administration among the youth surveys might lead to differences in their estimates. In addition, the youth survey estimates have sampling error, as demonstrated by the confidence intervals about the estimates from the comparison surveys. Questions and their orderings also differ among the surveys, as described in Appendix A, as do the modes of administration. The PATH Study, NHANES, and NSDUH use ACASI for the questions about tobacco use by youth, and these are administered individually in a household or mobile examination center setting. The NYTS is a pencil-and-paper survey administered in the classroom. Currivan et al. (2004) found that even when telephone ACASI was used, estimates of youth smoking prevalence were lower for a telephone survey of youth smoking than for a school-based survey of the same population (see also Fowler and Stringfellow, 2001, for a discussion of higher smoking rates in school-based surveys). In addition, school-based surveys often include students who are older than 17, which is the upper age limit for youth in the PATH Study.

The PATH Study's estimates of youth smoking are lower than comparable estimates from NHANES and NSDUH. Part of this difference may be sampling error and part may be attributable to differences among the survey wordings and administrations. Moreover, the comparison surveys are from different time periods. According to SAMHSA (2013, 2014), cigarette smoking among teens is dropping (from 2012 to 2013, the percentage of youth who had ever tried smoking dropped by 0.8 percentage points among 12-13 year olds, 1.6 percentage points among 14-15 year olds, and 2.7 percentage points among 16-17 year olds, with similar decreases from 2011 to 2012). The lower percentages found by the PATH Study may reflect, in part, a continuation of this trend. However, some of the differences among the estimates of youth smoking prevalence may be attributable to nonresponse bias or measurement error on the part of one or more of the surveys.

2.3 Statistical Approach for Addressing Nonresponse

2.3.1 Interviews

The primary approach for addressing nonresponse is to use differential weight adjustments. These adjustments are done at the household level and at the person level. The weight adjustments calibrate the estimates of demographic characteristics such as age, race, and sex to values calculated from the 1-year 2013 ACS (which are considered to be highly accurate because of the large sample size and high response rate for the ACS). They also correct for disparities among other characteristics that might be associated with the characteristics involved in the weighting

adjustments. 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.

Household Nonresponse-Adjusted Weights

Household IPS weights were computed for all sampled addresses in Wave 1. The IPS weights for responding households were adjusted to compensate for the estimated number of nonresponding households that were eligible for the PATH Study but did not complete the Household Screener. An eligibility adjustment was computed separately for each census region. Further adjustments were made within weighting classes based on information available for both responding and nonresponding households, namely the segments and blocks in which they are located. Census 2010 data were used to form weighting classes according to the percentage of occupied housing units, the percentage of population that is Black,¹⁵ the percentage of population that is Hispanic, and other information related to demographics and income. Census region and the urbanicity of the PSU and segment were also used when forming the weighting classes.

Within a weighting class, the IPS weights for the responding households were inflated proportionately to produce the same sum as the sum of the combined IPS weights of the responding and nonresponding households. The nonresponse-adjusted household weight is

$$HHNRWT_{ijk} = HHIPSWT_{ijk} \times \frac{\text{sum of HHIPSWT for eligible sampled households in weighting class}}{\text{sum of HHIPSWT for responding households in weighting class}}$$

The nonresponse-adjusted weights were raked to the 1-year 2013 ACS household counts by census region and household composition. Household composition was defined by the number of non-adult persons in the household (0, 1, or 2+) and the number of adult household members (1, 2, 3+). For raking purposes, the household composition was imputed for households missing this information using logical imputation.¹⁶ The final raked household weight is

$$HHRKWT_{ijk} = HHNRWT_{ijk} \times (\text{raking adjustment}).$$

¹⁵ Black is defined as Black alone, or in combination with other races.

¹⁶ See Lohr (2010) for a brief description of raking and imputation methods.

Person Nonresponse-Adjusted Weights

The raked household-level weight was used as the foundation for calculating the nonresponse-adjusted person-level weights, for both youth and adults. The initial person-level nonresponse-adjusted weight was computed as the product of the Household Screener raked weight $HHRKWT$ and the reciprocal of the within-household probability of selection for person l within household k of PSU i and segment j , as shown in the following formulas:

$$AP1BWT_{ijkl} = HHRKWT_{ijk} \times \frac{1}{\text{Probability adult } l \text{ selected at Phase 1 from household } (ijk)},$$

$$YBWT_{ijkl} = HHRKWT_{ijk} \times \frac{1}{\text{Probability youth } l \text{ selected from household } (ijk)}.$$

The probability of selection differed for adults and youth, as described in Section 2.2.1. Although shadow youth were not interviewed at Wave 1, a base weight was calculated for the shadow youth to serve as their base weight once they age up to the youth cohort:

$$SYBWT_{ijkl} = HHRKWT_{ijk} \times \frac{1}{\text{Probability shadow youth } l \text{ selected from household } (ijk)}.$$

Similar to the adjustment for Household Screener nonresponse, a nonresponse adjustment was performed to account for nonrespondents to the Youth Extended Interview. The weighting classes were formed using information similar to that used for the household-level nonresponse adjustment, and other variables from the Household Screener: age and sex of the household informant; count of adults in the household (0, 1, 2, 3, 4, 5+); and age, sex, and race/ethnicity of the youth. The youth weight adjusted for nonresponse is

$$YNRWT_{ijkl} = YBWT_{ijkl} \times \frac{\text{sum of YBWT for eligible sampled youth in weighting class}}{\text{sum of YBWT for responding youth in weighting class}}$$

For youth, the nonresponse-adjusted weights were raked to population totals from the 1-year 2013 ACS, using census region, age, race/ethnicity, and sex as raking variables. If variables used for nonresponse and/or raking adjustments were missing, they were imputed from the Household Screener or by logical or hot-deck imputation. After raking, the final weights for youth are denoted as $YRKWT$ and were computed as

$$YRKWT_{ijkl} = YNRWT_{ijkl} \times (\text{raking adjustment}).$$

A similar procedure was used to create raked weights SYRKWT for shadow youth at Wave 1.

Final weights for adults were computed in three steps. First, a nonresponse adjustment was performed using the information described above for the household-level nonresponse adjustment (with the exception of the segment urbanicity) and other variables from the Household Screener: count of adults in the household (0, 1, 2, 3, 4, 5+); and age, race/ethnicity, sex, and tobacco use status of the adult. The resulting adult weight for respondents to the Phase 2 Screener after adjusting for nonresponse between Phases 1 and 2 of the adult sampling procedure is

$$AP1NRWT_{ijkl} = AP1BWT_{ijkl} \times \frac{\text{sum of AP1BWT for adults sampled at Phase 1 in weighting class}}{\text{sum of AP1BWT for adults responding to Phase 2 Screener in weighting class}}$$

Second, the probability of selection at Phase 2 was used to calculate the Phase 2 weight:

$$AP2WT_{ijkl} = AP1NRWT_{ijkl} \times \frac{1}{\text{Probability adult } l \text{ from household } (ijk) \text{ selected at Phase 2}}$$

Finally, the Phase 2 adult weights were raked to independent population totals based on data from the 1-year 2013 ACS. The raking was done using combinations of census region, age, race/ethnicity, sex, and educational attainment. If variables used for nonresponse and/or raking adjustments were missing, they were imputed from the Household Screener or by logical or hot-deck imputation. The final raked weight is

$$ARKWT_{ijkl} = AP2WT_{ijkl} \times (\text{raking adjustment}).$$

Estimates calculated using the raked weights for variables of interest in the PATH Study are shown in Tables 2-16 through 2-27, which repeat the analyses used for Tables 2-5 through 2-12, 2-14, and 2-15. The estimates calculated using IPS weights are retained in these tables to facilitate comparisons of the estimates obtained using the two sets of weights. Confidence intervals are given for each of the IPS-weighted and raked-weighted estimates in each of Tables 2-16 through 2-27. The confidence intervals reported for the IPS-weighted estimates are the same as those given in the corresponding tables in Section 2.2.

Table 2-16. Race by age distribution, based on household enumeration

Race and age classification	Unweighted count	Weighted percentage, using household IPS weights	Confidence interval using household IPS weights	Weighted percentage, using household raked weights	Confidence interval using household raked weights	Percentage from ACS PUMS
Black* 18-24	3,669	2.3%	[2.0%, 2.6%]	2.0%	[1.8%, 2.3%]	2.0%
Black* 25+	17,934	11.4%	[9.8%, 12.9%]	11.1%	[9.8%, 12.5%]	10.5%
Non-Black 18-24	18,058	11.4%	[10.8%, 12.0%]	10.2%	[9.7%, 10.7%]	10.1%
Non-Black 25+	116,957	74.9%	[73.3%, 76.5%]	76.7%	[75.2%, 78.1%]	77.5%
Missing age or race	4,003					
Total	160,621	100.0%		100.0%		100.0%

*Black alone or in combination with other race(s).

Table 2-17. Distribution of male and female adults listed in the household enumeration

Sex	Unweighted count	Weighted percentage for adults, using household IPS weights	Confidence interval using household IPS weights	Weighted percentage for adults, using household raked weights	Confidence interval using household raked weights	Percentage from ACS PUMS
Male	77,088	48.0%	[47.8%, 48.3%]	47.7%	[47.5%, 47.9%]	48.2%
Female	83,500	52.0%	[51.7%, 52.2%]	52.3%	[52.1%, 52.5%]	51.8%
Missing	33					
Total	160,621	100.0%		100.0%		100.0%

The household raked weight HHRKWT adjusts the weights so that they agree with the 1-year 2013 ACS household counts by region and household size. They would therefore not be expected to bring person-level percentages of specific demographic groups closer to the ACS values. Tables 2-16 and 2-17 compare the estimated percentage of adults in the PATH Study household rosters to the ACS values for each race/age and sex group using the raked weights. The estimated percentages using the raked weights are not significantly different from the ACS values for race/age, although they are significantly different from the ACS values for sex. Table 2-21 indicates that using the raked adult weights produces distributions for race/age and sex that are practically identical to those from the ACS.

Tables 2-18 through 2-20 provide estimated distributions for household size and numbers of adults and youth per household, respectively; as expected, the raked weights align the estimated percentages with the 1-year 2013 ACS values. Confidence intervals shown in Tables 2-18 and 2-19 for the estimates computed with the raked weights are narrow because the raking constrains the estimates to accord with the ACS on the raking dimensions.

Table 2-18. Distribution of household size based on households responding to the Household Screener

Number of persons in household	Un-weighted count	Weighted percentage, using household IPS weights	Confidence interval using household IPS weights	Weighted percentage, using household raked weights	Confidence interval using household raked weights	Percentage from ACS PUMS
1	17,822	22.7%	[22.1%, 23.4%]	27.7%	[27.7%, 27.7%]	27.7%
2	26,098	33.0%	[32.5%, 33.5%]	33.7%	[33.7%, 33.7%]	33.7%
3	13,503	17.0%	[16.7%, 17.3%]	15.2%	[15.1%, 15.4%]	15.7%
4	11,720	14.7%	[14.3%, 15.1%]	12.9%	[12.7%, 13.1%]	13.0%
5+	10,055	12.6%	[12.0%, 13.2%]	10.5%	[10.3%, 10.6%]	9.9%
Total	79,198	100.0%		100.0%		100.0%

Table 2-19. Distribution of number of adults based on households responding to the Household Screener

Number of adults in household	Un-weighted count	Weighted percentage, using household IPS weights	Confidence interval using household IPS weights	Weighted percentage, using household raked weights	Confidence interval using household raked weights	Percentage from ACS PUMS
0-1*	21,869	27.8%	[27.1%, 28.5%]	33.3%	[33.3%, 33.4%]	33.3%
2	40,961	51.7%	[51.2%, 52.2%]	50.7%	[50.7%, 50.7%]	50.7%
3+	16,368	20.5%	[19.8%, 21.1%]	16.0%	[15.9%, 16.0%]	16.0%
Total	79,198	100.0%		100.0%		100.0%

*A small number of households contain only emancipated youth and hence contribute to the zero part of this category.

Table 2-20. Distribution of number of youth ages 12-17 based on households responding to the Household Screener

Number of youth ages 12-17 in household	Un-weighted count	Weighted percentage, using household IPS weights	Confidence interval using household IPS weights	Weighted percentage, using household raked weights	Confidence interval using household raked weights	Percentage from ACS PUMS
0	65,661	83.0%	[82.4%, 83.5%]	84.5%	[84.3%, 84.8%]	84.4%
1	9,569	12.0%	[11.7%, 12.4%]	11.0%	[10.8%, 11.2%]	11.1%
2+	3,968	5.0%	[4.7%, 5.2%]	4.4%	[4.3%, 4.6%]	4.5%
Total	79,198	100.0%		100.0%		100.0%

Tables 2-21 and 2-22 present estimates of demographic characteristics, education, and health insurance based on adult respondents in Wave 1, using the adult raked weight ARKWT. Raking adjusted the weights to match ACS totals for combinations of sex, age, race, ethnicity, and education categories. Raking increased the estimated percentage of adults with health insurance, however, as

noted in Section 2.2, the ACS estimates for the percentage of adults with health insurance were based on data collected before the launch of the Affordable Care Act.

Table 2-21. Demographic distributions based on adults responding to the Adult Extended Interview

	Adult respondents to Adult Extended Interview					ACS PUMS Percentage
	Unweighted count	Weighted percentage using adult IPS weights	Confidence Interval using adult IPS weights	Weighted percentage using adult raked weights	Confidence interval using adult raked weights	
Sex						
Male	16,309	45.7%	[44.9%, 46.6%]	48.1%	[48.1%, 48.1%]	48.1%
Female	15,982	54.3%	[53.4%, 55.1%]	51.9%	[51.9%, 51.9%]	51.9%
Missing	29					
Total	32,320	100.0%		100.0%		100.0%
Age group						
18-24	9,110	13.8%	[13.2%, 14.4%]	13.0%	[13.0%, 13.0%]	13.0%
25-44	11,280	35.8%	[35.0%, 36.6%]	34.3%	[34.3%, 34.4%]	34.3%
45-64	8,818	33.7%	[32.8%, 34.5%]	34.5%	[34.5%, 34.5%]	34.5%
65+	3,110	16.7%	[16.0%, 17.4%]	18.2%	[18.2%, 18.2%]	18.2%
Missing	2					
Total	32,320	100.0%		100.0%		100.0%
Race						
Black alone or in combination	5,538	14.1%	[12.1%, 16.1%]	13.1%	[12.9%, 13.3%]	12.5%
White alone	23,242	77.4%	[75.1%, 79.7%]	77.9%	[77.6%, 78.1%]	75.7%
Other	2,726	8.5%	[7.5%, 9.5%]	9.0%	[8.8%, 9.3%]	11.8%
Missing	814					
Total	32,320	100.0%		100.0%		100.0%
Ethnicity						
Hispanic	5,536	17.8%	[15.3%, 20.3%]	15.2%	[15.1%, 15.3%]	15.0%
Non-Hispanic	26,288	82.2%	[79.7%, 84.7%]	84.8%	[84.7%, 84.9%]	85.0%
Missing	496					
Total	32,320	100.0%		100.0%		100.0%

Table 2-22. Comparison of education level and health insurance status based on adults responding to the Adult Extended Interview

	Adult respondents to Adult Extended Interview					ACS PUMS Percentage
	Unweighted count	Weighted percentage using adult IPS weights	Confidence interval using adult IPS weights	Weighted percentage using adult raked weights	Confidence interval using adult raked weights	
Education						
18-24						
< HS, HS or GED	4,465	49.9%	[48.1%, 51.7%]	43.5%	[43.4%, 43.6%]	43.6%
> HS	4,593	50.1%	[48.3%, 51.9%]	56.5%	[56.4%, 56.6%]	56.4%
Subtotal	9,058	100.0%		100.0%		100.0%
25+						
< HS or GED	4,769	18.0%	[16.9%, 19.1%]	16.9%	[16.8%, 17.0%]	16.9%
HS	4,763	20.0%	[19.1%, 21.0%]	23.8%	[23.7%, 23.9%]	23.8%
Some college, no degree	7,650	31.3%	[30.2%, 32.3%]	29.2%	[29.1%, 29.3%]	29.2%
Bachelor degree	3,642	18.2%	[17.1%, 19.2%]	18.7%	[18.7%, 18.8%]	18.7%
> Bachelor degree	2,228	12.5%	[11.5%, 13.5%]	11.4%	[11.3%, 11.4%]	11.3%
Subtotal	23,052	100.0%		100.0%		100.0%
Missing	210					
Total	32,320					
Health insurance						
Yes	25,760	84.4%	[83.5%, 85.2%]	85.5%	[84.9%, 86.0%]	83.3%
No	6,156	15.6%	[14.8%, 16.5%]	14.5%	[14.0%, 15.1%]	16.7%
Missing	404					
Total	32,320	100.0%		100.0%		100.0%

Estimates of smoking prevalence in Table 2-23 using the raked weight ARKWT are similar to the estimates using the IPS weight AIPSWT; both are in the range of values obtained by other surveys. The use of raked weights resulted in a slight decrease in estimated smoking prevalence for females and non-white non-Hispanics.

Tables 2-24 and 2-25 examine the effect of the raked weight YRKWT on estimates calculated for youth. As expected, raking corrects for the overrepresentation of Hispanics among youth in Wave 1. The IPS-weighted estimates for youth age and sex agree with the 1-year 2013 ACS estimates. Consequently, raking had little if any effect on estimates of these characteristics. Cigarette smoking prevalence estimates with IPS weights and with raked weights were generally lower than estimates from other surveys although, as noted above, the surveys took place in different time periods.

Table 2-23. Current cigarette smoking based on adults responding to the Adult Extended Interview

	Sample size	PATH Study: Unweighted percentage	PATH Study: Weighted percentage, using adult IPS weights [95% confidence interval]	PATH Study: Weighted percentage, using adult raked weights [95% confidence interval]	Percentage from 2010-2011 TUS-CPS [95% confidence interval]	Percentage from 2013 NHIS [95% confidence interval]	Percentage from 2011-2012 NHANES* [95% confidence interval]	Percentage from 2013 NSDUH, original definition** [95% confidence interval]	Percentage from 2013 NSDUH, modified definition [95% confidence interval]
Current smoker	32,245	35.4%	18.6% [17.7%, 19.4%]	18.2% [17.7%, 18.7%]	16.1% [15.8%, 16.3%]	17.8% [17.2%, 18.4%]	19.8% [17.5%, 22.1%]	22.9% [22.6%, 23.2%]	21.0% [20.4%, 21.7%]
Current smoker, male	16,265	36.8%	21.2% [20.3%, 22.2%]	21.0% [20.3%, 21.6%]	18.0% [17.7%, 18.4%]	20.5% [19.5%, 21.4%]	23.9% [20.7%, 27.1%]	25.6% [25.1%, 26.1%]	23.8% [22.7%, 24.9%]
Current smoker, female	15,952	33.9%	16.4% [15.4%, 17.3%]	15.7% [15.0%, 16.3%]	14.2% [13.9%, 14.5%]	15.3% [14.6%, 16.0%]	16.0% [13.5%, 18.5%]	20.4% [20.0%, 20.8%]	18.4% [17.6%, 19.3%]
Current smoker, age 18-24	9,099	27.2%	20.1% [18.8%, 21.5%]	19.7% [18.5%, 20.9%]	17.1% [16.4%, 17.8%]	18.7% [16.8%, 20.5%]	20.4%*** [13.7%, 27.1%]	NA****	NA
Current smoker, age 25-44	11,260	41.2%	22.9% [21.8%, 24.1%]	23.1% [22.2%, 23.9%]	17.9% [17.5%, 18.4%]	20.1% [19.1%, 21.1%]	23.3% [20.0%, 26.7%]	NA	NA
Current smoker, age 45-64	8,784	40.6%	18.7% [17.7%, 19.8%]	18.4% [17.7%, 19.2%]	17.8% [17.4%, 18.2%]	20.0% [19.0%, 20.8%]	21.3% [18.3%, 24.2%]	NA	NA
Current smoker, age 65+	3,100	23.3%	7.7% [7.0%, 8.4%]	7.6% [7.0%, 8.3%]	7.8% [7.5%, 8.2%]	8.8% [8.0%, 9.6%]	9.2% [6.7%, 11.7%]	NA	NA
Current smoker, Hispanic	5,519	26.3%	13.4% [12.6%, 14.2%]	13.4% [12.6%, 14.2%]	10.9% [10.4%, 11.5%]	12.1% [11.0%, 13.2%]	16.6% [13.7%, 19.5%]	18.9% [18.1%, 19.7%]	15.2% [13.8%, 16.7%]
Current smoker, white non-Hispanic	19,268	38.8%	19.6% [18.%, 20.7%]	19.4% [18.6%, 20.1%]	17.5% [17.2%, 17.8%]	19.4% [18.5%, 20.2%]	20.2% [17.0%, 23.3%]	24.1% [23.7%, 24.5%]	22.8% [21.9%, 23.6%]
Current smoker, other non-Hispanic	6,904	33.3%	20.4% [19.3%, 21.5%]	18.3% [17.4%, 19.2%]	NA	16.7% [15.6%, 17.8%]	20.8% [16.6%, 24.9%]	21.8% [20.2%, 23.4%]	19.6% [18.1%, 21.2%]
Current every-day smoker	32,245	28.0%	14.8% [13.9%, 15.6%]	14.5% [14.0%, 15.0%]	12.7% [12.4%, 12.9%]	13.7% [13.1%, 14.2%]	16.4% [14.3%, 18.4%]	NA	NA
Current some-days smoker	32,245	7.4%	3.8% [3.6%, 4.1%]	3.7% [3.5%, 3.9%]	3.4% [3.3%, 3.5%]	4.1% [3.8%, 4.4%]	3.4% [2.7%, 4.1%]	NA	NA

*The smoking questions asked in NHANES for adults ages 20 and older differ from the questions asked for persons ages 12-19. The modes of administration also differ for the two age groups. The NHANES estimates presented in this table are for adults ages 20 and older.

**NSDUH's definition of a current cigarette smoker is someone who has smoked part or all of a cigarette in the past 30 days, which is more expansive than the definition used in the other surveys. However, NSDUH contains questions on lifetime smoking and current smoking. The modified definition uses these questions to construct a measure of "current smoking" that is comparable to that of the other surveys (Ryan et al., 2012). The construction of this variable is described in Appendix A. The estimates and confidence intervals for the NSDUH "original definition" (except for the "current smoker, other non-Hispanic" estimate) are from the published tables (SAMHSA, 2014); the estimates and confidence intervals for the "modified definition" are calculated from the public use data set. The estimate of current smoking for the "other non-Hispanic" group was not available from the published tables and it was also calculated from the public use data set.

***The estimate is for adults 20-24 years old.

****Detailed age information was not available in the public use file for NSDUH 2013.

Table 2-24. Demographic distributions based on youth ages 12-17 who completed the Youth Interview

	Youth ages 12-17 respondents to Youth Interview					ACS PUMS Percentage
	Unweighted count	Weighted percentage, using youth IPS weights	Confidence interval using youth IPS weights	Weighted percentage, using youth raked weights	Confidence interval using youth raked weights	
Sex						
Male	6,971	51.2%	[50.3%, 52.1%]	51.3%	[51.3%, 51.4%]	51.3%
Female	6,641	48.8%	[47.9%, 49.7%]	48.7%	[48.6%, 48.7%]	48.7%
Missing	39					
Total	13,651	100.0%		100.0%		100.0%
Age group						
12-13	4,684	34.3%	[33.5%, 35.2%]	33.7%	[33.7%, 33.7%]	33.7%
14-17	8,965	65.7%	[64.8%, 66.5%]	66.3%	[66.3%, 66.3%]	66.3%
Missing	2					
Total	13,651	100.0%		100.0%		100.0%
Race/ethnicity						
Hispanic	3,880	28.6%	[25.2%, 32.1%]	22.1%	[22.0%, 22.1%]	22.3%
Non-Hispanic white alone	6,616	48.4%	[45.1%, 51.7%]	54.6%	[54.5%, 54.7%]	54.5%
Non-Hispanic other	3,135	23.0%	[20.5%, 25.5%]	23.3%	[23.3%, 23.4%]	23.2%
Missing	20					
Total	13,651	100.0%		100.0%		100.0%

Table 2-25. Cigarette smoking* based on youth ages 12-17 who completed the Youth Interview

	Sample size	PATH Study: Unweighted percentage	PATH Study: Weighted percentage, using youth IPS weights [95% confidence interval]	PATH Study: Weighted percentage, using youth raked weights [95% confidence interval]	Percentage from 2011-2012 NHANES [95% confidence interval]	Percentage from 2013 NSDUH [95% confidence interval]	Percentage from 2012 NYTS [95% confidence interval]
Ever tried cigarette smoking, even one or two puffs	13,631	13.5%	13.5% [12.6%, 14.5%]	13.4% [12.6%, 14.2%]	20.5% [17.5%, 23.6%]	15.7% [15.4%, 16.0%]	25.6% [23.6%, 27.6%]
Ever tried smoking, male	6,959	13.9%	14.0% [13.0%, 15.1%]	14.0% [13.0%, 15.1%]	21.1% [15.9%, 26.3%]	16.3% [15.8%, 16.8%]	27.2% [25.0%, 29.3%]
Ever tried smoking, female	6,634	13.1%	13.1% [12.0%, 14.2%]	12.8% [11.8%, 13.8%]	20.0% [14.6%, 25.5%]	15.1% [14.6%, 15.6%]	24.0% [21.8%, 26.2%]
Ever tried smoking, age 12-13	4,675	4.6%	4.6% [3.9%, 5.4%]	4.5% [3.7%, 5.2%]	5.6% [1.9%, 9.4%]	4.0% [3.7%, 4.3%]	11.8% [10.2%, 13.4%]
Ever tried smoking, age 14-17	8,954	18.1%	18.2% [17.1%, 19.4%]	18.0% [17.0%, 19.0%]	28.3% [23.5%, 33.0%]	21.0% [20.0%, 22.0%]	32.5% [30.0%, 34.9%]
Have smoked in past 30 days	13,613	4.7%	4.7% [4.2%, 5.1%]	4.6% [4.2%, 5.0%]	6.9% [4.0%, 9.8%]	5.6% [5.4%, 5.8%]	8.7% [7.7%, 9.8%]

*Defined as ever tried a cigarette, even one or two puffs. For comparison, an additional measure of current smoking commonly applied to youth (having smoked at all in the past 30 days) is also included in this table.

2.3.2 Biospecimens

The adult raked weight ARKWT is also used for the analysis of adults in the full Wave 1 sample who provide biospecimens. Tables 2-26 and 2-27 present estimates of demographic characteristics, education, and health insurance for adults who provided biospecimens, using the adult raked weight ARKWT. Although the raking was performed on the adults responding to the Adult Extended Interview, and no additional adjustments were performed on the adults from whom biospecimens were collected, this raking brings the estimated sex, age, race, and ethnicity distributions for adults who provided urine specimens closer to estimates from the 1-year 2013 ACS. The raking also resulted in estimated percentages of females, Blacks, and Hispanics from the adults who provided blood samples that are closer to the 1-year 2013 ACS estimates, although the estimated age distribution for the adults who provided blood samples differs from the ACS distribution. Table 2-27 shows that the raking decreased the estimated percentages of adults at the lowest level of education, bringing those closer to the ACS estimates. Raking increased the estimated percentages of adults with health insurance for each type of biospecimen.

Table 2-28 gives estimates of current cigarette smoking prevalence for the adults from whom urine or blood specimens were collected. The estimates of cigarette smoking prevalence calculated with the raked weights are little changed from the estimates calculated with the IPS weights. The estimates of cigarette smoking prevalence are higher for both sets of biospecimen providers than for the full set of respondents to the Adult Extended Interview; however, all estimates are within the range of estimates from the comparison surveys.

The results in Tables 2-26 to 2-28 show estimates from the adults who provided urine or blood specimens, but not all specimens collected will be analyzed in the laboratory. Approximately 6,000 blood samples and 10,000 urine samples will be chosen initially for laboratory analysis. These biospecimens come from a probability sample of adults who are in specified tobacco use categories.¹⁷ If desired, an additional set of nonresponse-adjusted weights can be developed for these adults.

¹⁷ These categories are: current exclusive established users of cigarettes, current established users of other tobacco products, current experimental users only of any tobacco product, former established users of any product whose last use was within the past 12 months, and never users. In other categories of tobacco use, no samples are selected for laboratory analysis. Consequently, the samples of biospecimens will be representative of the persons in the specified tobacco use categories, but not of the adult population as a whole.

Table 2-26. Demographic distributions based on adults from whom urine or blood specimens were collected

	Adults from whom urine specimen is collected					Adults from whom blood specimen is collected					ACS PUMS Percentage
	Un-weighted count	Weighted percentage using adult IPS weights	Confidence interval using adult IPS weights	Weighted percentage using adult raked weights	Confidence interval using adult raked weights	Un-weighted count	Weighted percentage using adult IPS weights	Confidence interval using adult IPS weights	Weighted percentage using adult raked weights	Confidence interval using adult raked weights	
Sex											
Male	10,763	44.9%	[43.9%, 45.9%]	47.1%	[46.5%, 47.8%]	6,919	43.7%	[42.6%, 44.8%]	46.0%	[45.2%, 46.9%]	48.1%
Female	11,025	55.1%	[54.1%, 56.1%]	52.9%	[52.2%, 53.5%]	7,593	56.3%	[55.2%, 57.4%]	54.0%	[53.1%, 54.8%]	51.9%
Missing	13					6					
Total	21,801	100.0%		100.0%		14,518	100.0%		100.0%		100.0%
Age group											
18-24	6,457	14.9%	[14.2%, 15.6%]	14.1%	[13.8%, 14.4%]	3,884	13.3%	[12.5%, 14.0%]	12.5%	[12.0%, 12.9%]	13.0%
25-44	7,744	37.0%	[36.0%, 38.0%]	35.5%	[34.9%, 36.1%]	5,005	34.1%	[33.1%, 35.2%]	32.7%	[31.9%, 33.4%]	34.3%
45-64	5,725	33.0%	[32.0%, 33.9%]	33.7%	[33.1%, 34.3%]	4,191	35.6%	[34.4%, 36.7%]	36.1%	[35.2%, 37.0%]	34.5%
65+	1,873	15.1%	[14.3%, 15.9%]	16.7%	[16.0%, 17.3%]	1,438	17.1%	[16.1%, 18.1%]	18.8%	[17.9%, 19.6%]	18.2%
Missing	2					0					
Total	21,801	100.0%		100.0%		14,518	100.0%		100.0%		100.0%
Race											
Black alone or in combination	3,911	14.8%	[12.6%, 17.0%]	13.8%	[13.3%, 14.3%]	2,381	13.2%	[11.2%, 15.2%]	12.2%	[11.6%, 12.9%]	12.5%
White alone	15,531	77.0%	[74.6%, 79.4%]	77.8%	[77.1%, 78.5%]	10,637	79.3%	[77.1%, 81.5%]	80.5%	[79.6%, 81.4%]	75.7%
Other	1,801	8.2%	[7.2%, 9.2%]	8.4%	[7.9%, 9.0%]	1,152	7.5%	[6.7%, 8.4%]	7.3%	[6.7%, 7.9%]	11.8%
Missing	558					348					
Total	21,801	100.0%		100.0%		14,518	100.0%		100.0%		100.0%
Ethnicity											
Hispanic	3,870	18.6%	[16.0%, 21.2%]	16.0%	[15.5%, 16.5%]	2,451	16.7%	[13.9%, 19.5%]	14.3%	[13.7%, 15.0%]	15.0%
Non-Hispanic	17,633	81.4%	[78.8%, 84.0%]	84.0%	[83.5%, 84.5%]	11,884	83.3%	[80.5%, 86.1%]	85.7%	[85.0%, 86.3%]	85.0%
Missing	298					183					
Total	21,801	100.0%		100.0%		14,518	100.0%		100.0%		100.0%

Table 2-27. Comparison of education level and health insurance status based on adults from whom urine or blood specimens were collected

	Adults from whom urine specimen collected					Adults from whom blood specimen collected					ACS PUMS Percentage
	Unweighted count	Weighted percentage using adult IPS weights	Confidence interval using adult IPS weights	Weighted percentage using adult raked weights	Confidence interval using adult raked weights	Unweighted count	Weighted percentage using adult IPS weights	Confidence interval using adult IPS weights	Weighted percentage using adult raked weights	Confidence interval using adult raked weights	
Education											
18-24											
<HS, HS or GED	3,234	50.4%	[48.3%, 52.5%]	44.3%	[43.4%, 45.2%]	1,948	50.2%	[47.7%, 52.6%]	43.7%	[42.1%, 45.3%]	43.6%
>HS	3,197	49.6%	[47.5%, 51.7%]	55.7%	[54.8%, 56.6%]	1,928	49.8%	[47.4%, 52.3%]	56.3%	[54.7%, 57.9%]	56.4%
Subtotal	6,431	100.0%		100.0%		3,876	100.0%		100.0%		100.0%
25+											
< HS or GED	3,425	19.3%	[18.1%, 20.6%]	18.4%	[17.9%, 18.9%]	2,405	19.0%	[17.6%, 20.5%]	18.3%	[17.5%, 19.1%]	16.9%
HS	3,106	19.8%	[18.6%, 20.9%]	23.5%	[22.7%, 24.2%]	2,071	19.0%	[17.7%, 20.2%]	22.6%	[21.5%, 23.6%]	23.8%
Some college, no degree	5,238	32.4%	[31.3%, 33.4%]	30.3%	[29.7%, 30.9%]	3,721	33.4%	[31.9%, 34.8%]	31.3%	[30.4%, 32.2%]	29.2%
Bachelor degree	2,139	16.7%	[15.6%, 17.9%]	17.3%	[16.8%, 17.9%]	1,454	16.7%	[15.3%, 18.1%]	17.2%	[16.3%, 18.1%]	18.7%
> Bachelor degree	1,369	11.8%	[10.8%, 12.8%]	10.5%	[10.1%, 10.9%]	956	11.9%	[10.8%, 13.1%]	10.7%	[10.1%, 11.3%]	11.3%
Subtotal	15,277	100.0%		100.0%		10,607	100.0%		100.0%		
Missing	93					37					
Total	21,801					14,520					
Health insurance											
Yes	17,306	83.9%	[83.0%, 84.8%]	84.9%	[84.2%, 85.6%]	11,686	85.0%	[84.0%, 86.0%]	86.0%	[85.1%, 86.8%]	83.3%
No	4,335	16.1%	[15.2%, 17.0%]	15.1%	[14.4%, 15.8%]	2,758	15.0%	[14.0%, 16.0%]	14.0%	[13.2%, 14.9%]	16.7%
Missing	160					76					
Total	21,801	100.0%		100.0%		14,520	100.0%		100%		100.0%

Table 2-28. Current cigarette smoking based on adults from whom biospecimens were collected

	Sample size	PATH Study: Weighted cigarette smoking prevalence, using adult IPS weights [95% confidence interval]	PATH Study: Weighted cigarette smoking prevalence, using adult raked weights [95% confidence interval]	Percentage from 2010-2011 TUS-CPS [95% confidence interval]	Percentage from 2013 NHIS [95% confidence interval]	Percentage from 2011-2012 NHANES [95% confidence interval]	Percentage from 2013 NSDUH, original definition* [95% confidence interval]	Percentage from 2013 NSDUH, modified definition [95% confidence interval]
Adult respondent to Adult Extended Interview	32,245	18.6% [17.7%, 19.4%]	18.2% [17.7%, 18.7%]	16.1% [15.8%, 16.3%]	17.8% [17.2%, 18.4%]	19.8% [17.5%, 22.1%]	22.9% [22.6%, 23.2%]	21.0% [20.4%, 21.7%]
Adults providing urine	21,757	20.9% [20.0%, 21.9%]	20.5% [19.8%, 21.2%]	16.1% [15.8%, 16.3%]	17.8% [17.2%, 18.4%]	19.8% [17.5%, 22.1%]	22.9% [22.6%, 23.2%]	21.0% [20.4%, 21.7%]
Adults providing blood	14,493	21.6% [20.5%, 22.7%]	21.1% [20.2%, 21.9%]	16.1% [15.8%, 16.3%]	17.8% [17.2%, 18.4%]	19.8% [17.5%, 22.1%]	22.9% [22.6%, 23.2%]	21.0% [20.4%, 21.7%]

*NSDUH's definition of a current cigarette smoker is someone who has smoked part or all of a cigarette in the past 30 days. However, NSDUH contains questions on lifetime smoking and current smoking. The modified definition uses these questions to construct a measure of "current smoking" that is comparable to that of the other surveys (Ryan et al., 2012). The construction of this variable is described in Appendix A.

These weights would multiply ARKWT by an adjustment calculated using weighting adjustment cells based on Wave 1 characteristics such as demographics, education, employment status, health insurance, and tobacco use. These weights would not produce estimates for the full target population of the PATH Study, however, but only for the union of the tobacco use categories from which biospecimens were selected for analysis.

2.4 Summary of Findings

Response Rates

As reported in Section 2.1, the weighted response rates¹⁸ for the PATH Study Household Screener and Adult Interview and the biospecimen collections in Wave 1 are lower than projected (see Table 2-29), and the weighted response rates for all collections are higher than the worst-case scenario rates for the full sample provided in Attachment 22. The weighted response rate for the PATH Study Youth Interview is higher than projected.

Table 2-29. Summary of PATH Study Wave 1 overall response rates

Collection	Unweighted response rate, based on full Wave 1 sample	Weighted response rate, based on full Wave 1 sample	Projected response rate*	Worst-case scenario response rate*
Household Screener	54.1%	54.0%	70%	39.7%
Adult Extended Interview	74.8%	74.0%	85%	58.1%
Youth Interview	78.2%	78.4%	75%	–
Urine	67.5%	63.6%	80%	49%
Blood	44.9%	43.0%	65%	39%

*Provided in the request to OMB for Wave 1 data and biospecimen collection.

The differential weighted response rates are modest for tobacco use status and demographic subgroups (see Tables 2-1, 2-2, 2-3, and 2-4.) The largest differential weighted response rate, 11.5 percentage points, is for the age of adults who provide urine samples, which suggests a heightened potential for nonresponse bias. Notably, the differential weighted response rates for blood collection, ranging from 3.2 percentage points for ethnicity to 5.8 percentage points for race, were more consistent with those of other PATH Study collections.

¹⁸ These response rates were weighted with inverse probability of selection weights.

Nonresponse Bias Analysis

Nonresponse bias analysis indicates that estimates of key demographic and tobacco use variables calculated from the PATH Study Wave 1 sample with the inverse probability of selection weights are comparable to those produced by national cross-sectional surveys. However, the completed household interviews from the Wave 1 sample appear to underrepresent single- and two-person households relative to the 1-year 2013 ACS counts. The estimated percentage of persons who are non-Black and 25 years of age or older, from the household rosters, is also smaller than the corresponding estimate from the ACS.

Estimated distributions of demographic characteristics for adults completing the Adult Extended Interview are similar to those from the 1-year 2013 ACS for race (except for persons in the “other race” category). Persons in the “other race” category are also underrepresented among the persons providing blood or urine specimens. The estimated percentages of adults who are Hispanic are similar to ACS values for adults who provided blood specimens, but Hispanics are overrepresented among adults who responded to the Adult Extended Interview and those who provided urine specimens. In addition, the estimated percentage of adults who are between 18 and 24 years old or between 25 and 44 years old is higher for the PATH Study than for the ACS for adult respondents as a whole and for those who provided urine specimens. Males are underrepresented among respondents to the Adult Extended Interview, and also among the persons who provided blood or urine specimens.

When compared to national cross-sectional surveys that measure tobacco use (TUS-CPS, NHIS, NHANES, and NSDUH), estimates of adult cigarette smoking from the PATH Study Wave 1 sample are roughly in the middle of the range of estimates on smoking. There is no indication of nonresponse bias with respect to this measure.

Estimates of demographic characteristics of youth in Wave 1 align with the 1-year 2013 ACS for most demographic characteristics. However, the estimated percentage of youth who are Hispanic youth from the PATH Study is significantly higher than the corresponding percentage estimated from the ACS. (The nonresponse weight adjustments correct for this difference.)

PATH Study estimates of the selected youth cigarette smoking measure from the full Wave 1 sample are at the low end of estimates in comparison with national cross-sectional surveys that measure tobacco use (NHANES, NSDUH, and NYTS). However, estimates from the comparison surveys

are from 2011 through 2013 while those from the PATH Study are from September 2013 through December 2014, and evidence suggests the use of traditional cigarettes is declining among youth. The difference among surveys on time period alone is not large enough to account for the different estimates; as indicated in Section 2.2.2, time period is one of a number of factors that may explain the different estimates.

Statistical Approach for Addressing Nonresponse

The approach used to reduce potential nonresponse bias in the PATH Study is to adjust the weights of respondents at the household, adult, and youth levels to account for nonrespondents. Results of applying this approach to the full Wave 1 sample indicate the nonresponse adjustments are successful for reducing the discrepancy between the PATH Study estimates and 1-year estimates from the 2013 ACS with respect to demographic characteristics. Raked weights used for adults responding to the Adult Extended Interview reduced differences between the PATH Study and ACS for adults providing biospecimens as well, for sex and ethnicity. The raking did not reduce differences in the age distributions for the persons providing blood specimens, however. If desired, an additional set of nonresponse-adjusted weights could be created for the set of persons for whom biospecimens are analyzed, using weighting adjustment cells with respect to adults' Wave 1 characteristics.

Estimates of adult cigarette smoking using the IPS weights (before nonresponse adjustment) are in line with estimates from other surveys; agreement in these estimates is preserved using the nonresponse-adjusted weights. Weighting adjustments for youth corrected for the slight overestimate of the percentage of Hispanics among youth in Wave 1 but had little effect on the other demographic characteristics (i.e., IPS-weighted estimates already agreed with the ACS values) and estimates of youth cigarette smoking.

Wave 2 of the PATH Study is at the approximate mid-point of data and biospecimen collection. This section discusses response rates achieved to date for Wave 2, nonresponse analysis, and the Study's planned statistical approach for addressing nonresponse in Wave 2.

3.1 Response Rates

This section summarizes the three types of response rate calculations used for Wave 2 of the PATH Study: the interim retention and recruitment rates for the interviews and the interim response rates for the biospecimen collections. Retention for Wave 2 refers to obtaining a completed Wave 2 questionnaire from: (1) persons who completed the Adult Extended Interview in Wave 1, called continuing adults; or (2) persons who completed the Youth Extended Interview in Wave 1 and who are age 17 or younger at Wave 2, called continuing youth. Recruitment for Wave 2 applies to those who have aged up, either as shadow youth who have turned age 12 and are eligible to participate in the Wave 2 Youth Interview (called aged-up youths), or as adults who participated as youth in Wave 1 and have turned age 18, thus becoming eligible to participate as adults in Wave 2 (called aged-up adults). Response rates for biospecimen collections refer to the percentages of persons providing biospecimens among those who are asked to provide biospecimens.

3.1.1 Retention Rates for Continuing Adults and Continuing Youth

As stated in Section 1, the PATH Study Wave 1 sample was divided among four replicate groups. Replicate group 1, which consisted of the addresses that were released to the field in September 2013, obtained responses from 5,951 adults and 2,698 youth, and parental consent for 1,413 shadow youth. Approximately 98 percent of those cases had been released to the field for Wave 2 as of April 22, 2015. Replicate group 1 roughly corresponds to the earliest set of follow-ups in Wave 2, given the effort to schedule Wave 2 interviews on or near the Wave 1 anniversary dates.

Adults from Wave 1 are asked to complete an Adult Extended Interview in Wave 2. Persons who completed the youth questionnaire at Wave 1 are aged-up adults if they have attained age 18 by the

date of the Wave 2 interview, and are continuing youth if they are age 17 or younger on the date of the interview. Nonrespondents and interim cases for Wave 2, however, do not have a Wave 2 interview date, so the following procedure was used to classify them as aged-up adults or continuing youth for this report. The PATH Study field procedures call for beginning data collection contacts for all members of a household one month prior to the first day of the anniversary month of the household member with the earliest Wave 1 interview date. The age classification date is defined to be four months after the beginning of data collection contacts. Nonrespondents and interim Wave 1 youth who are age 17 or younger on the age classification date are classified as continuing youth, and all others are classified as aged-up adults.¹⁹ A similar classification rule is used for persons who were shadow youth at Wave 1. The Wave 1 shadow youth who completed a Youth Interview are classified as aged-up youth, as are nonrespondent and interim Wave 1 shadow youth who attained age 12 on or before the age classification date.

Table 3-1 displays the status of replicate group 1 cases that had been released to the field as of April 22, 2015 for the four categories of continuing adults, aged-up adults, continuing youth, and aged-up youth. The finalized cases include respondents, finalized nonrespondents, and 36 continuing adults known to be deceased.

Table 3-1. Status of Wave 2 released cases from replicate group 1, as of April 22, 2015

Group	Status of cases					
	Finalized		Interim		Total	
	n	%	n	%	n	%
Continuing adults	4,388	75.3	1,437	24.7	5,825	100.0
Aged-up adults	382	78.4	105	21.6	487	100.0
Continuing youth	1,765	81.7	396	18.3	2,161	100.0
Aged-up youth	431	78.5	118	21.5	549	100.0

¹⁹ Under this rule, for retention rate calculation purposes, nonrespondent and interim youth from Wave 1 are assigned to the category (continuing youth or aged-up adult) that would result if they completed an interview on the age classification date. To explore the sensitivity of the response rate calculations to this definition, the response rates were also calculated under an alternative classification that pro-rated the nonrespondents and interim cases who were age 17 at the beginning of the data collection and age 18 on the age classification date, according to the percentage of respondents with those ages who completed the adult questionnaire. Four persons changed age group classification under the alternative rule, with a negligible impact on retention and recruitment rates.

Method

Consistent with the response rate calculation guidelines specified by the Office of Management and Budget (2006), final retention rates for Wave 2 will be calculated for adults as the ratio of the number of Wave 2 Adult Interview completed cases (or sufficient partials) to the number of cases eligible for the Wave 2 Adult Interview. A simplified formulation will be used for this report because the eligibility status of some interim cases is unknown. The simplified formulation corresponds to AAPOR RR1 (AAPOR, 2011), which treats all completed cases from Wave 1, with the exception of persons known to be deceased, as eligible for Wave 2. The projected retention rates are therefore slightly conservative because some ineligible persons are included in the denominator.

If all the Wave 2 cases were finalized, the RR1 retention rate for continuing adults would be calculated as (number of completes or sufficient partials at Wave 2)/(number of Wave 1 completed cases – number of deceased persons), where the denominator can equivalently be expressed as the sum of the respondents and finalized nonrespondents. Because the PATH Study Wave 2 data collection is ongoing, however, the formula must consider “nonfinalized” or interim status cases as well as finalized cases; in this sense, the retention and recruitment rates presented in the interim report are “predicted.” Therefore, in this report, the unweighted retention rate for continuing adults is calculated as

$$RRU_{CA} = (C_{CA} + \sum_{i=1}^{I_{CA}} \hat{p}_{i,CA}) / (C_{CA} + N_{CA} + I_{CA}),$$

where

- C_{CA} = number of Wave 2 completed cases or sufficient partials among continuing adults;
- N_{CA} = number of Wave 2 finalized nonrespondents among continuing adults;
- I_{CA} = number of Wave 2 interim cases among continuing adults; and
- $\hat{p}_{i,CA}$ = predicted probability of interim continuing adult i becoming a respondent.

For continuing youth, the denominator of the response rate is defined using the age classification date described earlier in this section, and the unweighted retention rate is calculated as

$$RRU_{CY} = (C_{CY} + \sum_{i=1}^{I_{CY}} \hat{p}_{i,CY}) / (C_{CY} + N_{CY} + I_{CY}),$$

where

- C_{CY} = number of Wave 2 completed cases or sufficient partials among continuing youth;

- N_{CY} = number of Wave 2 finalized nonrespondents among Wave 1 Youth Extended Interview completed cases who were age 17 or younger on the age classification date;
- I_{CY} = number of Wave 2 interim cases among Wave 1 Youth Extended Interview completed cases who were age 17 or younger on the age classification date; and
- $\hat{p}_{i,CY}$ = predicted probability of interim continuing youth i becoming a respondent.

The weighted response rates are computed similarly, with each count replaced by the sum of the Wave 1 raked weights (YRKWT or ARKWT) for individuals in that category. The sums of the predicted probabilities for interim cases are replaced by $\sum_{i=1}^{I_{CA}} ARKWT_i \hat{p}_{i,CA}$ or $\sum_{i=1}^{I_{CY}} YRKWT_i \hat{p}_{i,CY}$, as appropriate.

The probability that an interim case will become a Wave 2 respondent is estimated using logistic regression, which is commonly used to predict response propensities. Models were fit to the sets of Wave 2 finalized and interim cases, separately for Wave 1 adults, Wave 1 youth, and Wave 1 shadow youth, to predict the probability of an interim case becoming a respondent as a function of respondent characteristics at Wave 1 (age, sex, race/ethnicity, tobacco use status²⁰), Wave 1 paradata (ever-refused at the household, parent, or person level in Wave 1, number of contact attempts in Wave 1, number of days in field in Wave 1), and Wave 2 paradata (ever had an interim adult or interim parent refusal). The number of weeks in field was used to divide the cases into nine time periods (weeks 1, 2, 3, 4-6, 7-9, 10-12, 13-15, 16-18, 19+). Logistic regression was used to estimate the probability that a case would respond in period k given that the case did not previously respond through a discrete time logistic survival model (Heeringa et al., 2010, Chapter 10; Berglund, 2011; Olson and Groves, 2012), and the response propensity for an interim case in time period k was estimated as $P(\text{respond in period } k + 1) + P(\text{respond in period } k + 2) + \dots + P(\text{respond in period } 9)$, where these probabilities were calculated as functions of the predictions from the logistic regression model.²¹

²⁰ Tobacco use status at Wave 1 was not available for the shadow youth and was omitted from the shadow youth model. Imputed age classification variables were used for persons missing age. Persons missing tobacco status were considered to be nonusers for purposes of predicting response propensities: this assumption had a negligible effect on the estimated response rates.

²¹ The results in the table were calculated using the conservative assumption that an interim case in time period k would not respond within time period k . Other models were also fit to investigate the sensitivity of the results to the model assumptions. These included logistic regression models (Groves et al., 2008; Wagner, 2010) that predicted response status from time in field, interim refusal status, demographic characteristics, and Wave 1 paradata: predictions from these models varied depending on whether interim cases were included in the model-fitting. The estimated unweighted retention rates from all models fit ranged between 82 percent and 90 percent for continuing adults.

Tables 3-2 and 3-3 provide predicted retention rates for continuing adults and continuing youth. In addition to the overall row, each table includes rows on tobacco use status, age, sex, race, and ethnicity subgroups based on self-reported data from the Wave 1 Extended Interviews. Persons with missing values for these characteristics from Wave 1 are excluded from the response rate calculation for the particular characteristic.

Results

The weighted predicted retention rates are approximately 84 percent for continuing adults and 89 percent for continuing youth. The unweighted predicted retention rates are approximately 85 percent for continuing adults and 90 percent for continuing youth. These approximately equal the projected retention rates of 85 percent for continuing adults and 90 percent for continuing youth provided in the Revision Request approved by OMB on 9/8/2014 for Wave 2. The predicted retention rates are sensitive to the models used for predicting the response propensity among the interim cases (approximately 25 percent of the continuing adults and 18 percent of the continuing youth), and the estimates of retention rates will be more accurate as more information accrues.

The variability among predicted retention rates for subgroups is small. For continuing adults, females appear to have slightly higher predicted retention rates than males, Blacks and Hispanics have higher predicted retention rates than non-Blacks and non-Hispanics, and persons ages 65 and older have lower predicted retention rates than younger persons. But these apparent differences depend largely on the disposition of the interim cases and no definitive conclusions can be made. The predicted retention rates appear similar for all the subgroups of continuing youth.

3.1.2 Recruitment Rates for Aged-up Adults and Aged-up Youth

This section reports recruitment rates for aged-up adults, who completed the Youth Interview at Wave 1 and are eligible for the Adult Interview at Wave 2, and aged-up youth, who were shadow youth at Wave 1 and are eligible for the Youth Interview at Wave 2. The Wave 2 Youth Interview is the first interview for aged-up youth, and aged-up adults are completing the Adult Interview for the first time. Table 3-1 displays the status of released cases for the aged-up adults and aged-up youth as of April 22, 2015.

Table 3-2. PATH Study Wave 2 predicted retention rates by Wave 1 characteristics: Adult Interview (continuing adults)

Characteristic ^a	A: Wave 2 Adult Interviews completed (n)	B: Wave 2 Interim likely to be completed ^b (n)	C: Wave 2 Finalized nonresponse (n)	D: Interim cases (n)	Unweighted predicted retention rate for Wave 2 ^c (%)	Weighted predicted retention rate for Wave 2 ^c (%)
Overall	4,038	903	314	1,437	85.4%	84.3%
Tobacco use status ^{de}						
Current user	1,791	426	152	662	85.1%	84.2%
Current non-user	2,136	449	148	731	85.7%	84.9%
Age						
18-24	1,091	334	56	521	85.4%	85.2%
25-44	1,379	350	102	533	85.8%	85.3%
45-64	1,172	179	93	302	86.2%	85.1%
65+	396	39	63	81	80.6%	79.9%
Sex ^e						
Male	1,991	492	168	783	84.4%	83.5%
Female	2,044	408	146	651	86.3%	84.9%
Race ^e						
White alone	2,915	600	254	986	84.6%	83.8%
Black alone or in combination	662	163	34	232	88.9%	87.6%
Other	357	112	18	171	85.9%	84.1%
Ethnicity ^e						
Hispanic	680	192	29	289	87.4%	87.7%
Non-Hispanic	3,298	691	280	1,119	84.9%	83.8%

^a The characteristics are as reported in the Adult Extended Interview at Wave 1.

^b Interim likely to be completed is the sum of the predicted probabilities of an interim case becoming a respondent over all interim cases.

^c Predicted retention rate = (A+B)/(A+C+D).

^d A tobacco user is defined as someone who currently uses one or more of the tobacco products covered by the PATH Adult Interview. A tobacco non-user is someone who does not currently use any of those tobacco products. A current user of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly. The products covered by the Adult Interview are cigarettes, traditional cigars, cigarillos, little filtered cigars, pipes, smokeless tobacco, hookah, e-cigarettes, and dissolvable tobacco.

^e The sum of counts for this category do not sum to the overall total due to missing values. The number of missing cases is 169 for tobacco use, 6 for sex, 160 for race, and 94 for ethnicity.

Table 3-3. PATH Study Wave 2 predicted retention rates by respondent characteristics: Youth Interview (continuing youth)

Characteristic ^a	A: Youth Interviews completed (n)	B: Interim likely to be completed ^b (n)	C: Finalized nonresponse ^c (n)	D: Interim cases (n)	Unweighted predicted retention rate for Wave 2 ^d (%)	Weighted predicted retention rate for Wave 2 ^d (%)
Overall	1,677	263	88	396	89.8%	89.2%
Tobacco use status						
Ever user	291	42	20	64	88.8%	87.9%
Never user	1,386	221	68	332	90.0%	89.5%
Age						
12-13	723	113	31	171	90.4%	89.9%
14-17	954	150	57	225	89.3%	88.7%
Sex ^f						
Male	870	132	44	196	90.3%	89.7%
Female	802	130	42	199	89.4%	88.8%
Race ^f						
White alone	1,108	152	71	237	89.0%	88.4%
Black alone or in combination	288	52	8	76	91.4%	91.2%
Other	197	36	4	49	93.2%	91.8%
Ethnicity ^f						
Hispanic	487	90	20	128	90.9%	90.4%
Non-Hispanic	1,158	168	65	263	89.2%	88.8%

^a The characteristics are as reported in the Youth Extended Interview at Wave 1.

^b Interim likely to be completed is the sum of the predicted probabilities of an interim case becoming a respondent over all interim cases.

^c Finalized nonresponse includes refused cases and all other nonresponding cases.

^d Predicted retention rate = (A+B)/(A+C+D).

^e A tobacco ‘ever user’ is someone who has ever used one or more of the tobacco products covered by the PATH Youth Interview. A tobacco ‘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 Youth Interview are cigarettes, traditional cigars, cigarillos, little filtered cigars, pipes, smokeless tobacco, hookah, e-cigarettes, dissolvable tobacco, bidis, and kreteks.

^f The sum of counts for this category do not sum to the overall total due to missing values. The number of missing cases is 8 for sex, 123 for race, and 40 for ethnicity.

Method

The methods described in Section 3.1.1 for estimating the retention rates were also used to estimate the recruitment rates for aged-up adults and aged-up youth. For aged-up adults, the unweighted recruitment rate is

$$RRU_{AUA} = (C_{AUA} + \sum_{i=1}^{I_{AUA}} \hat{p}_{i,AUA}) / (C_{AUA} + N_{AUA} + I_{AUA}),$$

where

- C_{AUA} = number of Wave 2 completed cases or sufficient partials among persons who completed the Youth Interview at Wave 1 and the Adult Interview at Wave 2;
- N_{AUA} = number of Wave 2 finalized nonrespondents among Wave 1 Youth Extended Interview completed cases who attained age 18 by the age classification date;
- I_{AUA} = number of Wave 2 interim cases among Wave 1 Youth Extended Interview completed cases who attained age 18 by the age classification date; and
- $\hat{p}_{i,AUA}$ = predicted probability of interim aged-up adult i becoming a respondent.

The unweighted recruitment rate for aged-up youth is estimated by

$$RRU_{AUY} = (C_{AUY} + \sum_{i=1}^{I_{AUY}} \hat{p}_{i,AUY}) / (C_{AUY} + N_{AUY} + I_{AUY}),$$

where

- C_{AUY} = number of Wave 2 completed cases or sufficient partials among persons who were shadow youth at Wave 1 and were administered the youth questionnaire at Wave 2;
- N_{AUY} = number of Wave 2 finalized nonrespondents among Wave 1 shadow youth who were age 12 by the age classification date;
- I_{AUY} = number of Wave 2 interim cases among Wave 1 shadow youth who were age 12 by the age classification date; and
- $\hat{p}_{i,AUY}$ = predicted probability of interim aged-up youth i becoming a respondent.

The weighted recruitment rates were calculated by substituting the sum of Wave 1 raked weights (YRKWT or SYRKWT) for individuals in each category for the counts, and replacing $\sum_{i=1}^{I_{AUA}} \hat{p}_{i,AUA}$, and $\sum_{i=1}^{I_{AUY}} \hat{p}_{i,AUY}$ by $\sum_{i=1}^{I_{AUA}} YRKWT_i \hat{p}_{i,AUA}$, and $\sum_{i=1}^{I_{AUY}} SYRKWT_i \hat{p}_{i,AUY}$ as appropriate.

Table 3-4 provides predicted recruitment rates for the Adult Interview for aged-up adults, and Table 3-5 provides predicted recruitment rates for the Youth Interview for aged-up youth. In addition to the overall row, each table includes rows on sex, race, and ethnicity subgroups; Table 3-4 also includes rows on tobacco use status. There are no rows corresponding to age subgroups in Table 3-4 or Table 3-5, because almost all of the aged-up adults are 18 years old and almost all of the aged-up youth are 12 years old; there are no rows for tobacco use status in Table 3-5, because no information was collected about the tobacco usage of shadow youth at Wave 1. Information from the Wave 1 Extended Youth Interview was used to define the demographic characteristics for the aged-up adults, and information from the Wave 1 Household Screener was used to define the

demographic characteristics for the aged-up youth. Persons with missing values for these characteristics on the Extended Youth Interview or Household Screener were excluded from the response rate calculation for that characteristic.

Table 3-4. PATH Study Wave 2 predicted recruitment rates by respondent characteristics: Adult Interview (aged-up adults)

Characteristic ^a	A: Adult Interviews completed (n)	B: Interim likely to be completed ^b (n)	C: Finalized nonresponse ^c (n)	D: Interim cases (n)	Unweighted predicted recruitment rate for Wave 2 ^d (%)	Weighted predicted recruitment rate for Wave 2 ^d (%)
Overall	365	56	17	105	86.4%	86.4%
Tobacco use status ^e						
Ever user	148	25	5	45	87.4%	87.3%
Never user	217	31	12	60	85.8%	85.7%
Sex						
Male	179	21	12	44	85.1%	85.4%
Female	186	35	5	61	87.7%	87.2%
Race/ethnicity ^f						
Non-Hispanic white alone	190	26	13	50	85.4%	85.5%
Other	173	30	4	55	87.5%	87.3%

^a The characteristics are as reported in the Youth Extended Interview at Wave 1.

^b Interim likely to be completed is the sum of the predicted probabilities of an interim case becoming a respondent over all interim cases.

^c Finalized nonresponse includes refused cases and all other nonresponding cases.

^d Predicted recruitment rate = (A+B)/(A+C+D).

^e A tobacco 'ever user' is someone who has ever used one or more of the tobacco products covered by the PATH Youth Interview. A tobacco '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 Youth Interview are cigarettes, traditional cigars, cigarillos, little filtered cigars, pipes, smokeless tobacco, hookah, e-cigarettes, dissolvable tobacco, bidis, and kreteks.

^f The sum of counts for this category do not sum to the overall total due to 2 missing values.

Table 3-5. PATH Study Wave 2 predicted recruitment rates by respondent characteristics: Youth Interview (aged-up youth)

Characteristic ^a	A: Youth Interviews completed (n)	B: Interim likely to be completed ^b (n)	C: Finalized nonresponse ^c (n)	D: Interim cases ^d (n)	Unweighted predicted recruitment rate for Wave 2 ^e (%)	Weighted predicted recruitment rate for Wave 2 ^e (%)
Overall	404	77	27	118	87.6%	87.5%
Sex						
Male	194	35	16	52	87.4%	86.7%
Female	210	42	11	66	87.8%	88.2%
Race/ethnicity ^f						
Non-Hispanic white alone	212	31	13	50	88.4%	88.2%
Other	191	46	14	68	86.8%	86.5%

^a The characteristics are as reported in the Household Screener at Wave 1.

^b Interim likely to be completes is the sum of the predicted probabilities of an interim case becoming a respondent overall all interim cases.

^c Finalized nonresponse includes refused cases and all other nonresponding cases.

^d Total interim includes ever refused interim cases and never refused interim cases.

^e Predicted recruitment rate = (A+B)/(A+C+D).

^f The sum of counts for this category do not sum to the overall total due to 1 missing value.

Results

The predicted recruitment rate is approximately 86 percent for aged-up adults, and variability in predicted recruitment rates is low among different subgroups of aged-up adults. This is slightly below the projected recruitment rate of 88 percent for aged-up adults in the Wave 2 Revision Request. The predicted recruitment rate for aged-up youth of 88 percent is higher than the projected rate of 85 percent in the Wave 2 Revision Request.

Subgroup recruitment rates in Table 3-4 for aged-up adults are similar, with no apparent patterns. There are also no apparent differences among recruitment rates for demographic subgroups of the aged-up youth (Table 3-5).

3.1.3 Biospecimen Collections

The PATH Study requests a urine sample from a subsample of continuing adults in Wave 2 who provided urine samples at Wave 1; it also requests urine and blood samples at Wave 2 from all aged-up adults. As of April 22, 2015, 1,839 continuing adults and 365 aged-up adults in replicate group 1 had completed the Wave 2 Adult Interview and had been asked to provide biospecimens.

Method

The response rates were calculated using the following formula:

$$\text{RRU} = \frac{\text{(Number of samples collected)}}{\text{(Number of adults from whom a sample is requested)}}$$

The urine response rate for continuing adults is based on the 1,839 adults who, as of April 22, 2015, were asked to provide a urine specimen following their Wave 2 Adult Interview. Similarly, the urine and blood response rates for aged-up adults are based on the 365 aged-up adults who completed the Wave 2 Adult Interview as of this date.

Tables 3-6 and 3-7 provide predicted unweighted and weighted response rates for the biospecimen collections. The table includes rows with response rates for tobacco use status, age, sex, race, and ethnicity subgroups. Information from the Wave 1 Adult Interview was used to define the tobacco use status and demographic characteristics for the continuing adults, and information from the Wave 1 Youth Interview was used to define the tobacco use status and demographic characteristics for the aged-up adults. Adults with missing values for a characteristic were excluded from the response rate calculation for that characteristic. Weights ARKWT were used to calculate the weighted response rates for Table 3-6, and weights YRKWT were used to calculate the weighted response rates for Table 3-7.

Results

The projected response rates for biospecimen collection in the Wave 2 Revision Request were 80 percent for urine collection among continuing adults and, among aged-up adults, 69 percent for the collection of urine and 45 percent for the collection of blood. To date, more than 96 percent of the

continuing adults asked to provide urine specimens have done so, exceeding the projected response rate. Among the aged-up adults, the weighted response rates for urine and blood collection are 82 percent and 43 percent, respectively; the response rate to date for urine collection exceeds the projected response rate, and the response rate to date for blood collection is slightly less than the rate that was projected. Subgroup differences in response rates are relatively small.

Table 3-6. PATH Study Wave 2 response rates by respondent characteristics: Urine collection (continuing adults)

Characteristic ^a	A: Adults requested to provide urine (n)	Urine		
		B: Urine collected (n)	Unweighted response rate for Wave 2 ^b (%)	Weighted response rate for Wave 2 ^b (%)
Overall	1,839	1,776	96.6%	96.1%
Tobacco use status ^{cd}				
User	1,170	1,130	96.6%	96.7%
Non-user	645	625	96.9%	95.8%
Age				
18-24	540	521	96.5%	95.3%
25-44	681	656	96.3%	96.7%
45-64	506	495	97.8%	96.2%
65+	112	104	92.9%	94.6%
Sex ^d				
Male	959	925	96.5%	96.6%
Female	879	850	96.7%	95.6%
Race ^d				
White alone	1,329	1,287	96.8%	95.8%
Black alone or in combination	295	283	95.9%	96.4%
Other	171	162	94.7%	97.1%
Ethnicity ^d				
Hispanic	325	311	95.7%	95.2%
Non-Hispanic	1,487	1,439	96.8%	96.3%

Note. This table presents results on a subsample of continuing adults (i.e., adults who completed an Adult Extended Interview in Wave 1) who provided a urine sample in Wave 1 and were asked to provide a sample in Wave 2.

^a The characteristics are as reported in the Wave 1 Adult Extended Interview.

^b Response rate = B/A.

^c A tobacco user is defined as someone who currently uses one or more of the tobacco products covered by the PATH Adult Interview. A tobacco non-user is someone who does not currently use any of those tobacco products. A current user of a given tobacco product is someone who currently uses the product every day or some days and: for cigarettes, has smoked at least 100 cigarettes in their lifetime and, for any other tobacco product, has reported they ever used that product regularly. The products covered by the Adult Interview are cigarettes, traditional cigars, cigarillos, little filtered cigars, pipes, smokeless tobacco, hookah, e-cigarettes, and dissolvable tobacco.

^d The sum of counts for this category do not sum to the overall total due to missing values. The number of missing cases is 24 for tobacco use, 1 for sex, 44 for race, and 27 for ethnicity.

Table 3-7. PATH Study Wave 2 response rates by respondent characteristics: Biospecimen collections (aged-up adults)

Characteristic ^a	A: Adult Interviews completed (n)	Urine			Blood		
		B: Collected (n)	Unweighted response rate for Wave 2 ^b (%)	Weighted response rate for Wave 2 ^b (%)	B: Collected (n)	Unweighted response rate for Wave 2 ^b (%)	Weighted response rate for Wave 2 ^b (%)
Overall	365	304	83.3%	82.3%	158	43.3%	43.2%
Tobacco use status ^c							
Ever user	148	129	87.2%	86.3%	70	47.3%	47.2%
Non-user	217	175	80.6%	79.8%	88	40.6%	40.7%
Sex							
Male	179	145	81.0%	80.8%	78	43.6%	44.1%
Female	186	159	85.5%	83.7%	80	43.0%	42.4%
Race ^d							
White alone	249	206	82.7%	82.5%	103	41.4%	42.6%
Black alone or in combination	57	50	87.7%	85.4%	26	45.6%	43.3%
Other	38	31	81.6%	78.3%	19	50.0%	44.0%
Ethnicity ^d							
Hispanic	96	83	86.5%	84.2%	44	45.8%	46.3%
Non-Hispanic	264	217	82.2%	81.9%	111	42.0%	42.1%

Note. This table presents results on aged-up adults (i.e., adults who completed a Youth Extended Interview in Wave 1 and became age-eligible for an Adult Interview in Wave 2).

^a The characteristics are as reported in the Youth Interview at Wave 1.

^b Response rate = B/A.

^c A tobacco ‘ever user’ is someone who has ever used one or more of the tobacco products covered by the PATH Youth Interview. A tobacco ‘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 Youth Interview are cigarettes, traditional cigars, cigarillos, little filtered cigars, pipes, smokeless tobacco, hookah, e-cigarettes, dissolvable tobacco, bidis, and kreteks.

^d The sum of counts for this category do not sum to the overall total due to missing values. The number of missing cases is 21 for race and 5 for ethnicity.

3.2 Nonresponse Bias Analysis

The standard approach for an analysis of nonresponse bias in a longitudinal cohort study such as the PATH Study would be to compare Wave 2 respondents with Wave 2 nonrespondents with respect to characteristics from Wave 1 (Bose and West, 2002; Javitz and Wagner, 2005; Brownstein et al., 2009). At the mid-point of Wave 2 data collection, there are a number of interim cases yet to be finalized as either respondents or nonrespondents. The number of finalized nonrespondents is small and does not include interim cases that will ultimately be nonrespondents. For this analysis, Wave 2

continuing adult respondents are compared with the finalized nonrespondents. To explore the sensitivity of results to the disposition of the interim cases, Wave 2 continuing adult respondents are also compared with *provisional nonrespondents*, defined to be the set of finalized nonrespondents plus interim refusals²² and persons who are difficult to locate. More than a third of the interim cases among provisional nonrespondents are expected to complete the Wave 2 interview; however, they are more likely to require intensive contact tracing and follow-up efforts than are other interim cases. For this reason, they are considered to be more similar to finalized nonrespondents than are other interim cases. For continuing youth and aged-up adults, the number of finalized nonrespondents is insufficient to permit meaningful comparisons (see Tables 3-3 and 3-4) and the respondents are compared only with the provisional nonrespondents.

Sections 2.2 and 2.3 described the weight construction for Wave 1 of the PATH Study. The raked weights from Wave 1 are designed to reduce the potential nonresponse bias from Wave 1. For Wave 2, the nonresponse bias analysis uses the raked weights from Wave 1: ARKWT for the continuing adults, YRKWT for the continuing youth and aged-up adults, and SYRKWT for the aged-up youth.

Tables 3-8 and 3-9 compare Wave 1 demographic characteristics and tobacco use rates for Wave 2 continuing adult respondents with the finalized nonrespondents and with the provisional nonrespondents. Table 3-10 presents similar comparisons for the persons who were interviewed as youth at Wave 1; the Wave 2 aged-up adults and continuing youth are combined for this analysis because the number of provisional nonrespondents among aged-up adults is too small to allow for meaningful comparisons).²³ The number of provisional nonrespondents is too small to permit comparisons of respondents and provisional nonrespondents for the aged-up youth, who do not self-report information in Wave 1. The recruitment rate calculations in Table 3-5, however, indicate that there are only small differences in response rates by sex and race/ethnicity.

²² Interim refusals for adults are cases that initially declined to participate in the Wave 2 interview but are being followed for refusal conversion attempts. Interim refusals for youth are those for whom the parent initially declined permission for the youth to participate in the Wave 2 interview, but are being followed for refusal conversion attempts.

²³ Because of the smaller sample sizes for the aged-up adults and continuing youth, the percentages using tobacco are not broken down by demographic subgroups.

Table 3-8. Comparison of Wave 2 Continuing Adult Interview respondents with finalized and provisional nonrespondents

Characteristic at Wave 1	Wave 2 respondents to Adult Interview		Wave 2 finalized nonrespondents to Adult Interview		Difference in weighted percentages [respondents – finalized nonrespondents]*	Wave 2 provisional nonrespondents to Adult Interview		Difference in weighted percentages [respondents – provisional nonrespondents]*
	Un-weighted count	Weighted percentage, using adult Wave 1 final weights	Un-weighted count	Weighted percentage, using adult Wave 1 final weights		Un-weighted count	Weighted percentage, using adult Wave 1 final weights	
		[95% confidence interval]		[95% confidence interval]			[95% confidence interval]	
Sex								
Male	1,991	47.2% [45.1%, 49.4%]	168	49.0% [41.2%, 56.8%]	-1.8% [-10.2%, 6.6%]	525	52.4% [48.0%, 56.8%]	-5.2% [-10.3%, 0.0%]
Female	2,044	52.8% [50.6%, 54.9%]	146	51.0% [43.2%, 58.8%]	1.8% [-6.6%, 10.2%]	418	47.6% [43.2%, 52.0%]	5.2% [0.0%, 10.3%]
Missing	3		0			1		
Age group								
Age 18-24	1,091	11.8% [11.0%, 12.7%]	56	7.0% [4.8%, 9.1%]	4.9% [2.7%, 7.1%]	296	14.1% [11.7%, 16.5%]	-2.3% [-4.8%, 0.2%]
Age 25-44	1,379	34.4% [32.5%, 36.3%]	102	27.9% [21.9%, 33.9%]	6.5% [0.3%, 12.8%]	323	33.5% [29.4%, 37.6%]	0.9% [-3.6%, 5.4%]
Age 45-64	1,172	35.5% [33.5%, 37.5%]	93	30.4% [23.7%, 37.1%]	5.1% [-2.2%, 12.5%]	226	30.3% [25.7%, 34.8%]	5.2% [0.0%, 10.5%]
Age 65+	396	18.3% [16.6%, 19.9%]	63	34.8% [27.2%, 42.4%]	-16.5% [-24.4%, -8.6%]	99	22.1% [17.4%, 26.8%]	-3.9% [-8.9%, 1.2%]
Missing	0		0			0		
Race/ethnicity								
Non-Hispanic white alone	2,438	66.2% [63.5%, 69.0%]	233	78.6% [72.3%, 84.9%]	-12.3% [-18.9%, -5.8%]	572	65.8% [61.0%, 70.6%]	0.4% [-4.1%, 5.0%]
Other	1,534	33.8% [31.0%, 36.5%]	76	21.4% [15.1%, 27.7%]	12.3% [5.8%, 18.9%]	348	34.2% [29.4%, 39.0%]	-0.4% [-5.0%, 4.1%]
Missing	66		5			24		
Health insurance								
Yes	3,202	85.9% [84.4%, 87.3%]	250	88.5% [84.0%, 93.0%]	-2.6% [-7.2%, 2.0%]	676	79.6% [76.2%, 83.0%]	6.3% [2.9%, 9.6%]
No	815	14.1% [12.7%, 15.6%]	54	11.5% [7.0%, 16.0%]	2.6% [-2.0%, 7.2%]	247	20.4% [17.0%, 23.8%]	-6.3% [-9.6%, -2.9%]
Missing	21		10			21		
Education								
< HS or GED	834	16.7% [15.0%, 18.3%]	66	22.6% [16.9%, 28.4%]	-6.0% [-12.1%, 0.2%]	207	22.2% [18.3%, 26.2%]	-5.6% [-9.8%, -1.4%]
HS	858	21.9% [20.2%, 23.7%]	68	25.1% [18.3%, 31.8%]	-3.1% [-9.9%, 3.7%]	221	24.6% [20.7%, 28.6%]	-2.7% [-6.7%, 1.4%]
Some college, no degree	1,434	31.6% [29.7%, 33.5%]	108	27.8% [21.9%, 33.6%]	3.8% [-2.5%, 10.1%]	327	30.7% [27.1%, 34.3%]	0.9% [-3.0%, 4.8%]
Bachelor degree +	899	29.8% [27.3%, 32.3%]	68	24.5% [17.1%, 32.0%]	5.3% [-2.6%, 13.1%]	181	22.4% [17.9%, 27.0%]	7.4% [2.3%, 12.4%]
Missing	13		4			8		

*Due to rounding, the difference in weighted percentages may differ by 0.1 from the results of [point estimate of respondents] – [point estimate of nonrespondents].

Table 3-9. Comparison of Wave 1 tobacco use rates for Wave 2 Continuing Adult Interview respondents with finalized and provisional nonrespondents

Characteristic at Wave 1	Wave 2 respondents to Adult Interview		Wave 2 finalized nonrespondents to Adult Interview		Difference in weighted percentages [respondents – finalized nonrespondents]* [95% confidence interval]	Wave 2 provisional nonrespondents to Adult Interview		Difference in weighted percentages [respondents – provisional nonrespondents]* [95% confidence interval]
	Unweighted count**	Weighted percentage, using adult Wave 1 final weights	Unweighted count	Weighted percentage, using adult Wave 1 final weights		Unweighted count	Weighted percentage, using adult Wave 1 final weights	
		[95% confidence interval]		[95% confidence interval]			[95% confidence interval]	
Overall	3,927	22.7% [21.4%, 24.1%]	300	24.5% [19.5%, 29.4%]	-1.7% [-6.9%, 3.4%]	911	28.8% [25.3%, 32.3%]	-6.0% [-9.7%, -2.4%]
Sex								
Male	1,936	27.2% [25.1%, 29.3%]	161	31.0% [22.9%, 39.0%]	-3.8% [-12.0%, 4.5%]	505	34.7% [29.7%, 39.6%]	-7.5% [-12.8%, -2.1%]
Female	1,988	18.8% [17.1%, 20.5%]	139	18.3% [11.8%, 24.8%]	0.5% [-6.1%, 7.1%]	405	22.5% [18.3%, 26.7%]	-3.7% [-7.8%, 0.5%]
Age group								
Age 18-24	1,070	29.1% [25.8%, 32.3%]	56	42.8% [29.6%, 55.9%]	-13.7% [-27.9%, 0.5%]	293	38.9% [32.6%, 45.3%]	-9.9% [-16.9%, -2.9%]
Age 25-44	1,351	26.6% [23.8%, 29.3%]	99	32.5% [23.3%, 41.7%]	-5.9% [-15.2%, 3.4%]	310	35.7% [30.0%, 41.4%]	-9.1% [-15.5%, -2.7%]
Age 45-64	1,124	23.3% [21.1%, 25.5%]	87	30.7% [19.7%, 41.8%]	-7.4% [-18.9%, 4.0%]	217	30.0% [23.7%, 36.3%]	-6.7% [-13.4%, -0.1%]
Age 65+	382	10.3% [7.8%, 12.7%]	58	8.5% [3.1%, 13.8%]	1.8% [-4.1%, 7.7%]	91	9.2% [4.5%, 13.9%]	1.1% [-4.0%, 6.2%]
Race/ethnicity								
Non-Hispanic white alone	2,382	23.6% [21.8%, 25.4%]	223	25.7% [19.7%, 31.8%]	-2.1% [-8.3%, 4.0%]	559	32.1% [27.0%, 37.2%]	-8.5% [-13.7%, -3.2%]
Other	1,487	20.9% [18.6%, 23.2%]	73	21.1% [12.0%, 30.2%]	-0.2% [-9.9%, 9.4%]	333	23.3% [18.9%, 27.6%]	-2.4% [-7.4%, 2.7%]

*Due to rounding, the difference in weighted percentages may differ by 0.1 from the results of [point estimate of respondents] – [point estimate of nonrespondents].

**The unweighted counts are for adults who provided information about tobacco use at Wave 1.

Table 3-10. Comparison of Wave 2 Aged-up Adult and Continuing Youth Interview respondents with provisional nonrespondents

Characteristic at Wave 1	Wave 2 respondents to Youth Interview		Wave 2 provisional nonrespondents to Youth Interview		Difference in weighted percentages [respondents – provisional nonrespondents]** [95% confidence interval]
	Unweighted count	Weighted percentage, using youth Wave 1 final weights [95% confidence interval]	Unweighted count	Weighted percentage, using youth Wave 1 final weights [95% confidence interval]	
Sex					
Male	1,049	51.4% [49.4%, 53.5%]	135	53.3% [46.9%, 59.8%]	-1.9% [-8.8%, 5.0%]
Female	988	48.6% [46.5%, 50.6%]	120	46.7% [40.2%, 53.1%]	1.9% [-5.0%, 8.8%]
Missing	5		2		
Age group					
Age 12-13	723	34.7% [32.6%, 36.8%]	74	28.8% [22.9%, 34.6%]	6.0% [-0.6%, 12.5%]
Age 14-17	1,319	65.3% [63.2%, 67.4%]	182	71.2% [65.4%, 77.1%]	-6.0% [-12.5%, 0.6%]
Missing	0		1		
Race/ethnicity					
Non-Hispanic white alone	996	56.1% [52.3%, 60.0%]	128	57.1% [49.0%, 65.1%]	-1.0% [-8.9%, 7.0%]
Other	1,006	43.9% [40.0%, 47.7%]	125	42.9% [34.9%, 51.0%]	1.0% [-7.0%, 8.9%]
Missing	40		4		
Tobacco use*					
Ever user	439	21.2% [18.9%, 23.6%]	60	23.6% [19.0%, 28.2%]	-2.3% [-7.4%, 2.7%]
Never user	1,603	78.8% [76.4%, 81.1%]	197	76.4% [71.8%, 81.0%]	2.3% [-2.7%, 7.4%]
Missing	0		0		

* A tobacco 'ever user' is someone who has ever used one or more of the tobacco products covered by the PATH Youth Interview. A tobacco '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 Youth Interview are cigarettes, traditional cigars, cigarillos, little filtered cigars, pipes, smokeless tobacco, hookah, e-cigarettes, dissolvable tobacco, bidis, and kreteks.

**Due to rounding, the difference in weighted percentages may differ by 0.1 from the results of [point estimate of respondents] – [point estimate of nonrespondents].

Among continuing adults, some trends differ for the comparisons with finalized nonrespondents and the comparisons with provisional nonrespondents. The estimated percentage of persons age 18-24 is higher for respondents than for finalized nonrespondents but the difference is not significant when the respondents and provisional nonrespondents are compared. The respondents have a higher rate of health insurance coverage than the provisional nonrespondents, but there is no significant difference in the rate of health insurance coverage between the respondents and final nonrespondents. The only demographic difference suggested by both sets of comparisons is for education. The estimated percentage of adults with less than a high school education is lower for respondents and the estimated percentage of adults with at least a college degree is higher for respondents, although the differences are not statistically significant between respondents and finalized nonrespondents. Respondents also have significantly lower tobacco use, overall and for most population subgroups, compared to provisional nonrespondents, but the differences are not significant when respondents are compared to finalized nonrespondents. Given that Wave 2 is underway and continuing to follow up and finalize interim cases, it is advisable to interpret these comparisons as preliminary and subject to change. With that said, the weighted predicted retention rates for Wave1 tobacco users and non-users from Table 3-2, which predicted the response propensity for all interim cases, are 84.2 percent and 84.9 percent, respectively.

The results in this report are based on preliminary data, and may change as more cases are finalized. If the trends seen among provisional nonrespondents continue as the interim cases are finalized, however, the PATH Study may experience attrition patterns that are similar to those in other longitudinal surveys. Thompson (2015) noted that younger persons and persons with lower educational levels are more difficult to retain in longitudinal surveys. Cunradi et al. (2005) and Young et al. (2006) have found that smokers were less likely to be retained in subsequent waves of surveys than nonsmokers.

Among Wave 1 youth, there are no significant differences between the respondents and the provisional nonrespondents.

3.3 Statistical Approach for Addressing Nonresponse

Initial weights for Wave 2 respondents will be adjusted to address nonresponse at Wave 2. Nonresponse adjustment cells will be formed using variables from Wave 1, including age, race, ethnicity, sex, employment status, education level, tobacco use, household composition, census

block characteristics, and the type of completed Wave 1 interview (Adult, Youth, or Household Screener).

Weight adjustments will be computed within cells formed from the cross-classification of variables available from Wave 1. Tree-based classification software will be employed to identify cells that distinguish between subgroups with different propensities to respond to the PATH Study (see Roth et al., 2006 and Schouten and deNoij, 2005). SAS macros will then be used to compute and apply the weighting adjustment factors and identify potential sources of concern in the adjustment process, such as small cell sizes and large adjustment factors.

The procedure described in Section 2.3.1 for Wave 1 can be used to address nonresponse for adults providing biospecimens at Wave 2.

3.4 Summary of Findings

Response Rates

Because the PATH Study Wave 2 data collection is ongoing, response rates²⁴ were calculated using predicted response propensities for interim cases, as described in Section 3.1. Table 3-11 indicates that the weighted predicted retention rates²⁵ for continuing adults and continuing youth are slightly lower than the projected rates, although the results are inconclusive because only 76 percent of the replicate group 1 cases have been finalized to date and the response status is predicted for the remaining 24 percent using statistical models. The weighted predicted recruitment rate for aged-up adults is slightly lower than the projection of 88 percent, and the weighted predicted recruitment rate for aged-up youth is higher than the projected value of 85 percent.

The response rate for each biospecimen collection is calculated as the percentage of persons who were requested to give a biospecimen who provided it. For both continuing adults and aged-up adults in Wave 2, the response rate for urine collection is higher than projected. Blood is collected only from aged-up adults in Wave 2, among whom the response rate for blood collection is slightly lower than projected.

²⁴ Response rates include retention rates for continuing adults and youth, recruitment rates for aged-up adults and youth, and response rates for providing biospecimens.

²⁵ These response rates were weighted using the raked weights ARKWT, YRKWT, and SYRKWT.

Table 3-11. Summary of PATH Study predicted response rates for Wave 2

Group	Unweighted predicted response rate	Weighted predicted response rate	Projected response rate*
Continuing adults, Adult Interview	85.4%	84.3%	85%
Continuing youth, Youth Interview	89.8%	89.2%	90%
Aged-up adults, Adult Interview	86.4%	86.4%	88%
Aged-up youth, Youth Interview	87.6%	87.5%	85%
Continuing adults, urine collection	96.6%	96.1%	80%
Aged-up adults, urine collection	83.3%	82.3%	69%
Aged-up adults, blood collection	43.3%	43.2%	45%

*Provided in the Revision Request to OMB for Wave 2 data and biospecimen collection.

The predicted response rates exhibit little variability across population subgroups. For continuing adults, predicted retention rates are slightly higher for females, Blacks, Hispanics, and adults under age 65. For continuing youth, aged-up adults, and aged-up youths, no evidence was found to indicate the retention or recruitment rates differ across population subgroups. Response rates for biospecimen collection are close to or exceed the projected rates for all categories.

Nonresponse Bias Analysis

The nonresponse bias analysis suggests that estimated percentages of persons with less than a high school degree tend to be lower for the Wave 2 respondents than for the Wave 2 finalized and provisional nonrespondents, although the difference between the respondents and finalized nonrespondents is not statistically significant. Tobacco use rates are not significantly different between respondents and finalized nonrespondents, although the provisional nonrespondents exhibit higher tobacco use rates than the respondents. Provisional nonrespondents also have lower health insurance coverage than respondents, although the difference between respondents and finalized nonrespondents is not significant.

Statistical Approach for Addressing Nonresponse

For Wave 2, weights of respondents will be adjusted to account for nonrespondents by forming weighting adjustment cells using Wave 1 characteristics of respondents and nonrespondents. Consequently, nonresponse-adjusted weights of Wave 2 respondents will sum to Wave 1 population totals. This weighting will compensate for differences between respondents and nonrespondents with respect to sex, age, other demographic variables, and Wave 1 tobacco use status.

Conclusions and Implications for Study Going Forward

4

This section summarizes the findings presented in this report on the PATH Study's Wave 1 and Wave 2 response rates, nonresponse bias analysis, and approach to addressing nonresponse. Its conclusions are based on the full sample for Wave 1 and on the data collected from replicate group 1 during the first 6 months (out of 12) of Wave 2. The section closes with a discussion of the implications of the conclusions for the study going forward.

Conclusions

Response rates in Wave 1 for the Household Screener and Adult Extended Interview were lower than projected in the Non-substantive Change Request for Wave 1 of the PATH Study but higher than the worst-case scenario. However, nonresponse bias analysis found that many characteristics of respondents in Wave 1 align with the 1-year estimates from the 2013 ACS. Exceptions were found for single-person households, education, and ethnicity when comparing PATH Study estimates using IPS weights to 1-year 2013 ACS estimates. Estimates of cigarette smoking among adults in Wave 1 are within the range of estimates found by other national health studies. Moreover, when full sample estimates were adjusted for nonresponse using the raked weights, they more closely approximated the ACS estimates, and adult smoking rates remained essentially the same.

The response rate for the Wave 1 Youth Interview was higher than projected. Nonresponse bias analysis among youth found that many characteristics of respondents were consistent with the 1-year estimates from the 2013 ACS, with the exception of ethnicity. When the full sample estimates were adjusted for nonresponse among youth, they more closely approximated the 2013 ACS estimates, but the ever-tried-smoking rates for youth remained lower than those found by other national studies.

The response rates for urine and blood collections in Wave 1 were lower than projected and exceeded the worst-case scenario response rates. Despite this, nonresponse bias analysis found that many of the characteristics of respondents were generally aligned with estimates of these characteristics from the 1-year 2013 ACS. In addition, when the sample estimates were adjusted for nonresponse, they were found to approximate the ACS estimates more closely.

The unweighted retention rates for Wave 2 continuing adults and continuing youth, calculated using predictions of response propensity for the interim cases, are close to those projected in the Revision Request for Wave 2; the weighted predicted retention rates are less than two percentage points lower than the projections. The estimated recruitment rate for aged-up adults is 1.6 percentage points lower than the projected rate, and the estimated recruitment rate for aged-up youth is approximately 2.5 percentage points higher than the projected rate. The Wave 2 response rates for biospecimens also approximately equal or exceed the projected rates. The estimated percentage of persons with at least a college degree is higher when calculated from the respondents than from the finalized or provisional nonrespondents; the estimated percentage of persons with less than a high school degree is lower. There is no evidence that tobacco use differs between respondents and finalized nonrespondents, but it is significantly lower among respondents than among provisional nonrespondents. There is also no evidence of nonresponse bias for the continuing youth, aged-up adults, or aged-up youth. However, as noted, it is important to regard these findings as preliminary pending finalization of interim cases and the rest of data collection in Wave 2.

Implications for the Study Going Forward

Findings on the response rates, nonresponse bias analysis, and approach to addressing nonresponse for Wave 1 and Wave 2 have three important implications for the PATH Study. First, the PATH Study should continue implementing new approaches to increase response rates for Wave 2 and subsequent waves. The lower than projected sample sizes resulting from the Wave 1 response rates underscore the need to achieve high response rates in each of the follow-up waves. The PATH Study is continually seeking ways to boost the response rates. For example, in Wave 2, it has enhanced its efforts to communicate by text message and email with participants who indicated they may be contacted in these ways. Participants continue to access information about the study on the participant pages of the PATH Study website, which also allows adult participants to update their contact information. The Study employs special interviewers with skills in refusal conversion and has a cadre of experienced traveling interviewers that help to augment staffing in specific areas. For Wave 3 and potential future waves, the PATH Study will vary the appearance of materials it provides participants to enhance their interest and engagement, and it will provide a certificate of appreciation to participating youth. In addition, the Study will take extra steps to interview continuing adults who have relocated to group quarters facilities since their initial interviews.

Second, data analyses will need to consider the smaller-than-expected sample sizes for Wave 1. Adjustments may be necessary, such as combining some subgroups in analyses. Future planning for longitudinal analyses will need to account for the declining rates of cigarette smoking in the U.S. population in general, notably among youth. For the PATH Study, the trends observed among youth mean smaller sample sizes for youth smokers and therefore less statistical power for examining within-person changes among youth cigarette smokers. At the same time, the larger sample size of youth nonsmokers provides more power for examining the initiation of cigarette smoking over time. The impact of the smaller sample size for youth smokers must also be considered in the context of the recent and rapid increase of youth use of alternative tobacco products such as e-cigarettes and hookah (CDC, 2015a).

In its review of this report, the Division of Data Policy of the Assistant Secretary for Planning and Evaluation (ASPE), U.S. Department of Health and Human Services (HHS) asked what is the minimum acceptable sample size for conducting a further wave and what happens if the sample size falls below this minimum. These questions can be answered in the context of minimum sample sizes required to detect meaningful differences in subgroups within or across data collection waves on key measures. For example, to detect a difference between Wave 1 and Wave 2 of less than 5 percentage points for the percent of youth who smoke menthol cigarettes, power analyses²⁶ for the Wave 2 Revision Request indicated that the expected sample sizes would be sufficient for 14-17 year old youth, but not for 12-13 year olds. Cases such as this can be dealt with by combining age subgroups. In such a case, the PATH Study would need to consider the current and projected sample sizes, rates of attrition, and whether to replenish the affected subgroup samples in a future data collection wave.

Third, on an ongoing and consistent basis, the PATH Study should examine the sample sizes achieved for Wave 1 and Wave 2 to date, as well as those projected for Wave 3. In this way, it would be possible to detect differential rates of attrition among subgroups early and make extra efforts to retain persons in subgroups of special analytic interest. If consideration is given to replenish the sample at some point, the replenishment could consider the subgroup sample sizes in the continuing cohort. This may be necessary, for example, for the shadow sample, which serves as a reservoir for aged-up youth in each wave but which will be exhausted after Wave 3.

²⁶ The power analyses assumed power equal to 0.8, two-sided tests of significance, and alpha equal to 0.05. These analyses made several additional assumptions, including those on the population prevalence of the behavior and the design effects for the data collections.

References

- Agaku, I.T., King, B.A., and Dube, S.R. (2014). Current cigarette smoking among adults – United States, 2005-2012. *Morbidity and Mortality Weekly Report*, 63, 29-34.
- American Association of Public Opinion Research (AAPOR). (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, 7th ed. http://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/StandardDefinitions2011_1.pdf.
- Berglund, P.A. (2011). An overview of survival analysis using complex survey data. *Proceedings of the 2011 SAS Global Forum*, <http://support.sas.com/resources/papers/proceedings11/338-2011.pdf>.
- 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.
- Brault, M.W. (2013). *Non-response bias in the 2013 CPS ASEC Content Test*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington DC, http://www.copafs.org/UserFiles/file/fcsm/H3_Brault_2013FCSM.pdf.
- 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). (2013). *National Survey on Drug Use and Health, 2012*. ICPSR34933-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-11-26. doi:10.3886/ICPSR34933.v1. Persistent URL: <http://doi.org/10.3886/ICPSR34933.v1>.
- Center for Behavioral Health Statistics and Quality (CBHSQ), Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services (HHS), and RTI International. (2014). *Results from the 2013 National Survey on Drug Use and Health: Detailed Tables. Prevalence Estimates, Standard Errors, P Values, and Sample Sizes*. Rockville, MD: Substance Abuse and Mental Health Services Administration, <http://www.samhsa.gov/data/sites/default/files/NSDUH-DefTabsPDFWHTML2013/Web/HTML/NSDUH-DefTabsTOC2013.htm>.
- Centers for Disease Control and Prevention (CDC). (2013). Tobacco product use among middle and high school students — United States, 2011 and 2012. *Morbidity and Mortality Weekly Report*, 62, 893–897. Retrieved from www.cdc.gov/mmwr/preview/mmwrhtml/mm6245a2.htm.

- Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). (2014a). *National Health and Nutrition Examination Survey Data, 2011-2012*. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, <http://www.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Questionnaire&CycleBeginYear=2011>.
- Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). (2014b). *National Health Interview Survey Data, 2013*. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, http://www.cdc.gov/nchs/nhis/nhis_2013_data_release.htm.
- Centers for Disease Control and Prevention (CDC). (2015a). *E-cigarette use triples among middle and high school students in just one year*. Press Release (April 16). Atlanta, GA: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, <http://www.cdc.gov/media/releases/2015/p0416-e-cigarette-use.html>.
- Centers for Disease Control and Prevention (CDC). (2015b). *National Youth Tobacco Survey Data, 2013*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, http://www.cdc.gov/tobacco/data_statistics/surveys/nyts.
- 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., Brick, J.M., Couper, M., Kalsbeek, W., Harris-Kojetin, B., Kreuter, F., Pennell, B., Raghunathan, T., Schouten, B., Smith, T., Tourangeau, R., Bowers, A., Jans, M., Kennedy, C., Levenstein, R., Olson, K., Peytcheva, E., Ziniel, S., and Wagner J. (2008). Issues facing the field: Alternative practical measures of representativeness of survey respondent pools. *Survey Practice*, 1(3).
- Heeringa, S.G., West, B.T., and Berglund, P.A. (2010). *Applied Survey Data Analysis*. Boca Raton, FL: CRC Press.
- 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.
- Lohr, S. (2010). *Sampling: Design and Analysis, 2nd ed.* Boston: Brooks/Cole.

- Office of Management and Budget. (2006). *Standards and Guidelines for Statistical Surveys*, available at http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf.
- Olson, K., and Groves, R.M. (2012). An examination of within-person variation in response propensity over the data collection field period. *Journal of Official Statistics*, 28, 29-51.
- Roth, S., Montaquila, J., and Chapman, C. (2006). Nonresponse bias in the 2005 National Household Education Surveys Program. Technical Report. (NCES 2007-016). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Ryan, H., Trosclair, A., and Gfroerer, J. (2012). Adult current smoking: Differences in definitions and prevalence estimates—NHIS and NSDUH, 2008. *Journal of Environmental and Public Health*, online.
- Särndal, C.-E., and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. Hoboken, NJ: Wiley.
- SAS Institute, Inc. (2011). *SAS/STAT® 9.3 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.
- Schouten, B., and de Nooij, G. (2005). *Nonresponse adjustment using classification trees*. Discussion paper 05001, Statistics Netherlands. Available at www.cbs.nl.
- Song, Y. (2013). *Rotation group bias in smoking prevalence estimates using TUS-CPS*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington DC, paper available at http://www.fcsm.gov/13papers/I3_Song_2013FCSM.pdf, and slides available at http://www.copafs.org/UserFiles/file/fcsm/I3_Song_2013FCSM.pdf.
- Substance Abuse and Mental Health Services Administration (SAMHSA). (2013). Substance Abuse and Mental Health Services Administration, *Results from the 2013 National Survey on Drug Use and Health: Summary of National Findings*, NSDUH Series H-48, HHS Publication No. (SMA) 14-4863. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2014. Retrieved from <http://www.samhsa.gov/data/sites/default/files/NSDUHresultsPDFWHTML2013/Wcb/NSDUHresults2013.htm>. The Survey Documentation and Analysis (SDA) system, National Survey on Drug Use and Health, 2013. <http://www.icpsr.umich.edu/cgi-bin/SDA/SAMHDA/hsda?samhda+35509-0001>.
- Thomson, M.E. (2015). Using longitudinal complex survey data. *Annual Review of Statistics and its Application*, 2, 305-320.
- United Nations. (2005). *Designing Household Survey Samples: Practical Guidelines*. United Nations Publication ST/ESA/STAT/SER.F/98, New York: United Nations. Available at <http://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf>.

- United States Census Bureau (2015). *American Community Survey 2013 ACS 1-year PUMS Files*. Retrieved from http://www.census.gov/acs/www/Downloads/data_documentation/pums/ACS2013_PUMS_README.pdf.
- United States Department of Commerce, Census Bureau. (2012). *National Cancer Institute-sponsored Tobacco Use Supplement to the Current Population Survey (2010-11)*: <http://appliedresearch.cancer.gov/studies/tus-cps/>.
Technical documentation: <http://www.census.gov/cps/methodology/techdocs.html>.
Retrieved from <http://appliedresearch.cancer.gov/studies/tus-cps/results/data1011/table1.html>.
- Wagner, J. (2010). The fraction of missing information as a monitoring tool for survey data quality. *Public Opinion Quarterly*, 74(2):223-243.
- Wang, K., Murphy, J., Baxter, R., and Aldworth, J. (2005). *Are two feet in the door better than one? Using process data to examine interviewer effort and nonresponse bias*. Paper presented at the 2005 Federal Committee on Statistical Methodology conference, http://www.fcsm.gov/05papers/Wang_Aldworth_etal_VIB.pdf.
- 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 on the PATH Study and Other Surveys

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

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

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

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

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

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

Table A-1. Question used to define “current smoking” in the PATH Study, TUS-CPS, NHIS, NHANES, and NSDUH

PATH Study	TUS-CPS**	NHIS	NHANES	NSDUH (original definition)	NSDUH (modified definition)*
Question to define current smoking (answers defining current smoking given in parentheses)					
“Have you ever smoked a cigarette, even one or two puffs?” (yes) and “Do you now smoke cigarettes every day, some days, or not at all?” (every day or some days) and “How many cigarettes have you smoked in your entire life? A pack usually has 20 cigarettes in it.” (100 or more cigarettes (5 packs or more))	“Have you smoked at least 100 cigarettes in your entire life?” (yes) and “Do you now smoke cigarettes every day, some days, or not at all?” (every day or some days) (PEA1, PEA3)	“Have you smoked at least 100 cigarettes in your ENTIRE LIFE?” (yes) and “Do you NOW smoke cigarettes every day, some days or not at all?” (every day or some days) (SMQEV, SMKNOW)	“{Have you/Has SP} smoked at least 100 cigarettes in {your/his/her} entire life?” (yes) and “{Do you/Does SP} now smoke cigarettes every day, some days or not at all?” (every day or some days) (SMQ020, SMQ040)	“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)	“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 “Have you smoked at least 100 cigarettes in your entire life?” (yes)
Age range included in estimate					
18+	18+	18+	20+	18+	18+
Exclusions from population					
Includes only civilian, non-institutionalized population. Excludes residents of group quarters, active military.	Includes only civilian, non-institutionalized population.	Includes only civilian noninstitutionalized population. Several segments of the population excluded, such as: patients in long-term care facilities; persons on active duty with the Armed Forces; persons incarcerated in the prison system; and U.S. nationals living in foreign countries.	Includes only civilian, non-institutionalized population.	Includes only civilian, non-institutionalized population. Excludes homeless persons who do not use shelters, military personnel on active duty, and residents of institutional group quarters.	
Proxy responses allowed					
No	Yes	Yes, for individuals physically or mentally incapable of responding.	No	No	No

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

** Proxies are allowed if 4th callback, the person will not return before closeout, or the household is getting irritated. See http://appliedresearch.cancer.gov/studies/tus-cps/surveys/tuscps_english_2010.pdf, p3.

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

PATH Study	NHANES	NSDUH	NYTS
Question to define ever tried cigarette smoking (answers defining ever tried cigarette smoking given in parentheses)			
“Have you ever tried cigarette smoking, even one or two puffs?” (yes)	“About how many cigarettes have you smoked in your entire life?” (SMQ621, values of 2-8 (more than a puff to 100 or more cigarettes)) I have never smoked, not even a puff (1), 1 or more puffs but never a whole cigarette (2), 1 cigarette (3), 16 to 25 cigarettes (6), 2 to 5 cigarettes (4), 26 to 99 cigarettes (7), 6 to 15 cigarettes (5), 100 or more cigarettes (8)	CG01 Have you ever smoked part or all of a cigarette? (yes)	Have you ever tried cigarette smoking, even one or two puffs? (Qn7 value of 1, Yes)
Questions for determining whether have smoked in past 30 days			
“Have you ever tried cigarette smoking, even one or two puffs?” (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)	“During the past 30 days, on how many days did you smoke cigarettes?” (SMQ640, Recoded to SMD641 in SMQ_G file, number of day smoked, values of 1 through 30)	CG05 [IF CG01 = 1 OR CGREF1 = 1] Now think about the past 30 days – that is, from [DATEFILL] up to and including today. During the past 30 days, have you smoked part or all of a cigarette?	During the past 30 days, on how many days did you smoke cigarettes? (Qn13 values of 2 through 7)
Ages of youth in survey			
12-17	12-17	12-17	12-17 year old students in public or private schools
Exclusions from population			
Residents of group quarters	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.	Only includes youth who attend either public or private schools.
Other comments			
	Those missing SMQ621 values are excluded from the estimates. Those with SMQ621=1, 2, 77 or 99 (never smoked, less than 1 cigarette, RF, DK) had SMD640 recoded to 0 (0 cigarette smoked in past 30 days) due to skip pattern.	The Center for Behavioral Health Statistics and Quality (2013, 2014) give estimates and the standard errors of the estimates.	The survey is administered by teachers in the classroom setting.