

**Attachment 21**

**PATH Study  
Interim Report on Baseline Data  
and Biospecimen Collection**

**June 26, 2014**

# Population Assessment of Tobacco and Health Study

## Interim Report to the Office of Management and Budget on Baseline Data and Biospecimen Collection

April 28, 2014

**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

Email: [PATHprojectofficer@mail.nih.gov](mailto:PATHprojectofficer@mail.nih.gov)

## Table of Contents

<u>Section</u>		<u>Page</u>
1	Introduction .....	1
	1.1 Overview of Sample Design for Baseline Wave.....	2
	1.2 Predictor Sample .....	3
2	Response Rates Associated with the Household Screener.....	5
	2.1 Method .....	5
	2.2 Results.....	7
3	Response Rates Associated with the Adult and Youth Interviews .....	9
	3.1 Adult Extended Interview .....	9
	3.2 Youth Interview .....	12
4	Response Rates Associated with the Biospecimen Collections .....	14
	4.1 Method .....	14
	4.2 Results.....	16
5	Nonresponse Bias Analysis .....	17
	5.1 Method .....	17
	5.2 Results.....	20
6	Statistical Approach for Addressing Nonresponse.....	33
	6.1 Computation of Nonresponse-Adjusted Weights .....	33
	6.2 Results.....	36
7	Discussion.....	46
	7.1 Summary of Findings .....	46
	7.2 Conclusions and Implications for Study Going Forward .....	49
	References.....	52

<b><u>Appendix</u></b>		<b><u>Page</u></b>
A	Cigarette Smoking Questions on the PATH Study and Other Surveys .....	A-1

**Table**

2-1	PATH Study baseline predicted response rates for the predictor sample, by address characteristics: Household Screener .....	8
3-1	PATH Study baseline predicted response rates for the predictor sample, by respondent characteristics: Adult Extended Interview .....	11
3-2	PATH Study baseline predicted response rates for the predictor sample, by respondent characteristics: Youth Interview .....	13
4-1	PATH Study baseline predicted response rates for the predictor sample, by respondent characteristics: Biospecimen collections .....	15
5-1	Race by age distribution, based on the household enumeration .....	21
5-2	Distribution of male and female adults listed in the household enumeration.....	21
5-3	Distribution of household size based on households responding to the Household Screener.....	22
5-4	Distribution of number of adults based on households responding to the Household Screener.....	22
5-5	Distribution of number of youth ages 12-17 based on households responding to the Household Screener.....	22
5-6	Demographic distributions based on adults responding to the Adult Extended Interview, and on adults from whom urine, buccal, and/or blood specimens were collected .....	24
5-7	Comparison of education level and health insurance status based on adults responding to the Adult Extended Interview, and on adults from whom urine, buccal, and/or blood specimens were collected .....	25

<b><u>Table</u></b>	<b><u>Page</u></b>	
5-8	Current cigarette smoking based on adults responding to the Adult Extended Interview.....	27
5-9	Current cigarette smoking based on adults from whom biospecimens were collected.....	29
5-10	Demographic distributions based on youth ages 12-17 who completed the Youth Interview.....	30
5-11	Cigarette smoking based on youth ages 12-17 who completed the Youth Interview.....	31
6-1	Race by age distribution, based on household enumeration.....	36
6-2	Distribution of male and female adults listed in the household enumeration.....	37
6-3	Distribution of household size based on households responding to the Household Screener.....	37
6-4	Distribution of number of adults based on households responding to the Household Screener.....	37
6-5	Distribution of number of youth ages 12-17 based on households responding to the Household Screener.....	38
6-6	Demographic distributions based on adults responding to the Adult Extended Interview, and on adults from whom urine, buccal, and/or blood specimens were collected.....	40
6-7	Comparison of education level and health insurance status based on adults responding to the Adult Extended Interview, and on adults from whom urine, buccal, and/or blood specimens were collected.....	41
6-8	Current cigarette smoking based on adults responding to the Adult Extended Interview.....	42
6-9	Current cigarette smoking based on adults from whom biospecimens were collected.....	43

<b><u>Table</u></b>		<b><u>Page</u></b>
6-10	Demographic distributions based on youth ages 12-17 who completed the Youth Interview.....	44
6-11	Cigarette smoking* based on youth ages 12-17 who completed the Youth Interview .....	45
7-1	Summary of PATH Study baseline overall response rates for the predictor sample .....	46
A-1.	Question used to define “current smoking” in the PATH Study, TUS-CPS, NHIS, NHANES, and NSDUH .....	A-2
A-2.	Questions used for youth cigarette smoking in the PATH Study, NHANES, NSDUH, and NYTS.....	A-4
 <b><u>Figure</u></b>		
1-1	Schematic of the PATH Study sample design with counts for the predictor sample .....	4

The Population Assessment of Tobacco and Health (PATH) Study is in the midst of completing the baseline wave of its planned 3-year data and biospecimen collection effort. The Office of Management and Budget (OMB) approved the PATH Study's non-substantive change request for the baseline wave collection on August 23, 2013 (0925-0664). The terms of clearance of OMB's approval state: "Before submitting the second wave of data collection to OMB for approval under the PRA (Paperwork Reduction Act), NIDA/FDA should report to OMB regarding the response rates associated with the baseline (screening, interview completion, and bio-specimen response), the results of nonresponse analysis, the statistical approach for addressing nonresponse, and the implications for the study going forward."

This report is submitted by NIDA/FDA to meet OMB's terms of clearance. The contents are presented as specified in the terms of clearance: Sections 2, 3, and 4 present the response rates; Section 5 provides the results of a nonresponse analysis; Section 6 discusses the statistical approach for addressing nonresponse; and Section 7 summarizes the findings and considers their implications. The rates provided in this report are for the "predictor sample," the probability sample of addresses selected for the main study that were released to field interviewers early in the field period as the first priority of field work. These rates for the predictor sample are compared throughout this report to corresponding rates projected for 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 the baseline wave of data and biospecimen collection.) The report covers approximately 5 months of the PATH Study's 12-month baseline, from September 12, 2013 to February 26, 2014, and the analyses are performed on data collected in a subsample of the full study sample called the predictor sample.

The next section provides an overview of the sample design for the PATH Study baseline wave and a description of the predictor sample on which this interim report is based. Information on the study background and overall design is provided in Supporting Statement A of the PATH Study's non-substantive change request for the baseline wave.

## 1.1 Overview of Sample Design for 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 also excluded, with the exception of college students. A four-stage stratified area probability sample design is used, with a two-phase design for sampling the adult cohort at the final stage. The sampling rates for adults vary 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 is a county or group of counties. For the second stage, within each selected PSU, smaller geographical segments were formed and then a sample of these segments was drawn. At the third stage, the sampling frame consists of the residential addresses in the U.S. The main source of these addresses is the Postal Service (USPS) Computerized Delivery Sequence Files (CDSFs).

The fourth stage selects persons from the sampled households. A roster of all the members in the sampled household is constructed by interviewing one adult household member (referred to as the household informant) to list the members and collect some information about each one for use in sampling the three 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.

Given the possible misreporting of tobacco use status of each adult in the household by the household informant, two-phase sampling is used for adult selection. The Phase 1 sampling depends on the age, race, and tobacco use information provided by the household informant. The Phase 2 sampling is based on the self-reported age, race, and tobacco use status, obtained by interviewing the individuals sampled at the first phase. The sampling rates for the two phases are 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 is selected using probability sampling methods, it is representative of the U.S. civilian non-institutionalized population 12 years of age and older.



## 1.2 Predictor Sample

Figure 1-1 is a graphic presentation of the sample design. It presents counts for the sampling stages/phases and data collection outcomes for the predictor sample. The PATH Study baseline 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 primary sampling unit (PSU). Each separate replicate group is therefore also representative of the civilian non-institutionalized U.S. population, because it is a probability sample from the set of segments in the frame. The data collection plan calls for the release of replicate groups to the field in a sequential manner (i.e., replicate group 1 in September 2013, replicate group 2 in November 2013, replicate group 3 in February 2014, and replicate group 4 in May 2014).

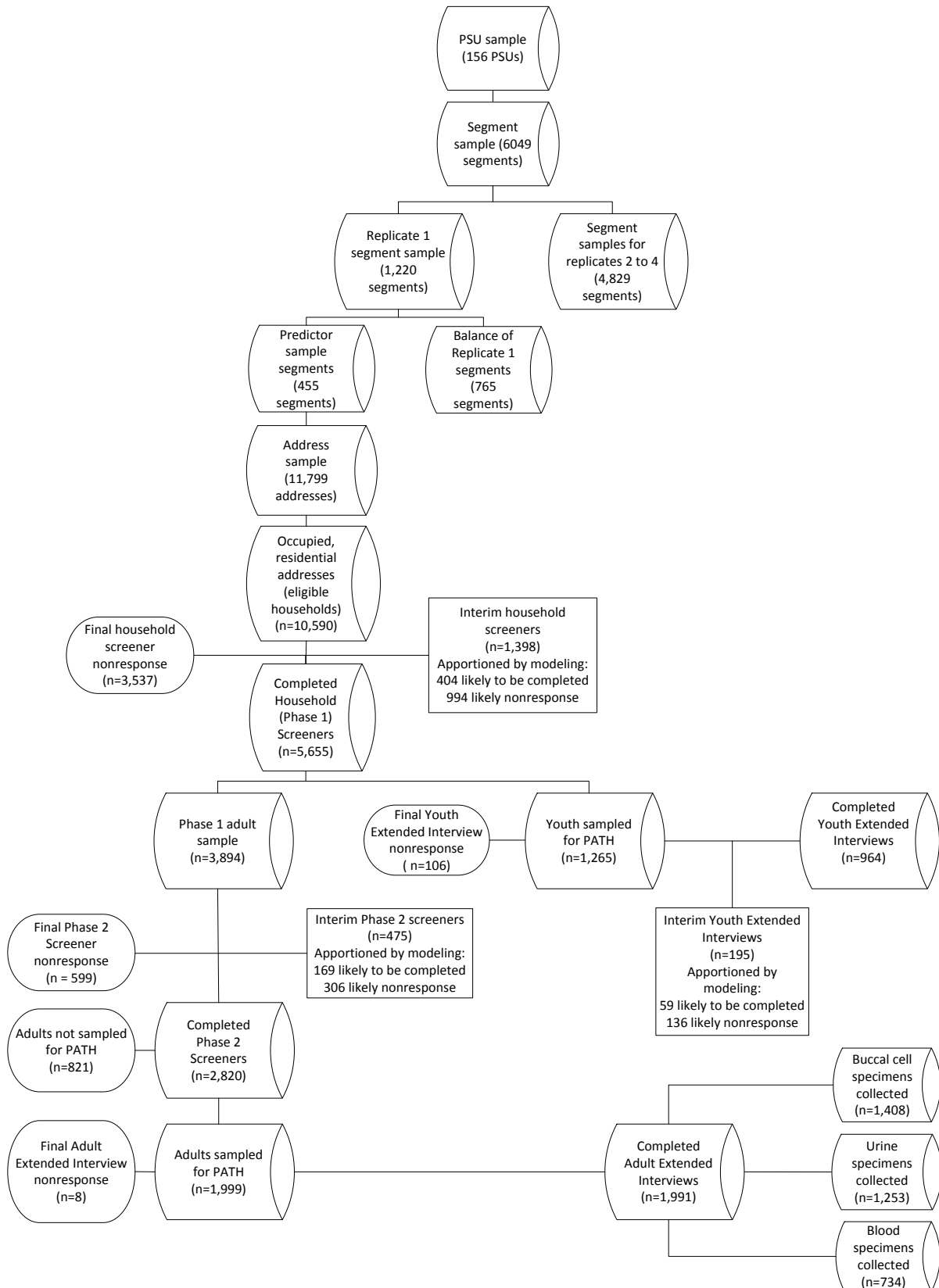
A random sample of  $r$  segments selected from each PSU in replicate group 1, where  $r$  is the nearest integer to  $(3/8) \times$  (number of segments in the PSU), was designated as the predictor sample. Because the predictor sample is a randomly selected subsample of the full sample, it is also approximately representative<sup>1</sup> of the civilian non-institutionalized U.S. population. Those segments were assigned early in the field period so that preliminary estimates from the predictor sample would be available to inform the response rates and nonresponse analysis in this report. The predictor sample consists of 455 segments selected from the 1,220 segments in replicate group 1, with representation from all 156 PSUs in the PATH Study sample; this sample includes 11,799 addresses, of which 10,590 addresses were eligible.

Weighting is discussed briefly in Sections 2.1, 3.1, and 4.1. More detailed information on it is provided in Sections 5 and 6.

---

<sup>1</sup> Some segments had incomplete information about addresses from the USPS CDSFs, and field staff listed the addresses in these segments using traditional, on-the-ground, in-person listing methods. The segments that were listed were not given a chance of selection for the predictor sample, because the extra time required for listing meant that survey data from such segments was not available within the required timeframe. A total of 59 out of the 1,220 segments in replicate group 1 were selected for listing. Similarly, the predictor sample addresses did not include addresses added to the sample as a result of the address verification (AV) procedure. The AV procedure was performed on a probability sample of the non-listed segments to ensure complete coverage for the PATH Study. In each segment selected for the AV procedure, field interviewers canvassed the segment and listed addresses not on the USPS CDSFs for potential inclusion in the sample, along with the addresses selected from the CDSF. The small number of additional listed addresses from the 50 predictor sample segments undergoing the AV procedure were not available in time to be included in the predictor sample. The predictor sample therefore does not have representation from the listed segments or addresses not found on the USPS CDSFs.

Figure 1-1. Schematic of the PATH Study sample design with counts for the predictor sample



# Response Rates Associated with the Household Screener

## 2

The baseline Household Screener (also referred to as “Phase 1” in Section 5) combines typical screener functions (e.g., enumerating the household, collecting basic demographic information about each member, collecting some household-level data, and selecting participants for the study) with a special purpose for the PATH Study, which is to collect minimum information on each adult’s tobacco use. This allows for classifying 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 conduct the Household Screener in person using computer-assisted personal interviewing (CAPI).

### 2.1 Method

As of February 26, 2014, 9,192 (86.8%) Household Screener cases were finalized, and 1,398 (13.2%) cases were still being followed up in field work.

Response rates presented in this report were computed in a manner consistent with the response rate formula specified 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.<sup>2</sup> 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:

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

---

<sup>2</sup> The predictor sample does not have any partial completes.

This response rate formula applies most directly to data collections that have been completed. Because the PATH Study baseline data and biospecimen collection is ongoing, however, the formula must consider nonfinalized or interim status cases as well as finalized cases; in this sense, the response rates presented in the interim report are “predicted.” Hence, the unweighted unit response rates used for this interim report are as follows:

$$RRU = ((C+(X * IR)+(Y * IO))/(C+R+NC+O+1(U)+IR+IO), \text{ where additionally}$$

- IR = number of ever refused interim cases;
- IO = number of never refused interim cases;
- X = probability of IR cases becoming respondents; and
- Y = probability of IO cases becoming respondents.

For this report, the PATH Study modeled the probabilities of interim cases becoming respondents, X and Y, using a procedure to project outcomes in biostatistics.<sup>3</sup> The mean probabilities of the interim cases are 0.11 for households that have ever refused, and 0.47 for the households that have never refused. These probabilities are consistent with the resolution rates found for interim cases based on number of call attempts by Wang et al. (2005). Note that the predicted response rates assume all pending cases are eligible (i.e., e = 1).

Table 2-1 provides overall predicted response rates for the Household Screener and response rates for subgroups of sampled households that belong to Census block groups with various characteristics. After the characteristic column, the table includes columns on the number of completed cases, number of interim cases likely to become respondents, number of finalized nonresponse cases, number of total interim cases, unweighted response rates, and weighted response rates. 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). Without weighting, the response rates would be expected to be biased. The Household Screener inverse probability of selection (IPS) weights were calculated as the inverse of the selection probabilities for all households sampled (responding households and nonresponding households).(See Section 5.1 for additional information on weighting.)

---

<sup>3</sup> The procedure entailed two steps. First, interim cases were divided into two categories: households that had refused one or more times (households that had ever refused), and households that had never refused. Second, for each category, a cumulative incidence model (Gooley et al., 1999) was fit to finalized cases. This model was used to estimate the probability that an interim case would become a respondent within 15 contact attempts as a function of the number of contact attempts to date. For the purpose of modeling the predictor sample, the results indicated up to 15 contact attempts captured the effects of the majority of field effort while still providing sufficient sample size of finalized cases from which to estimate probabilities of completion.

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 (>29.1% of persons ages 25 and older with Bachelor’s degrees) was 51.9 percent; it was 60.8 percent for addresses in Census block groups with “low” levels of education. Comparing subgroups of responding and nonresponding households on response rates informs an assessment of the extent to which the responding addresses represent all sampled addresses and, ultimately, the population of inference. To include information on the characteristics of both respondents and nonrespondents, subgroups are defined by the characteristics of the Census block groups in which the sampled addresses are located; this information is from the 5-year (2008 to 2012) American Community Survey (ACS).<sup>4</sup> The “high” and “low” subgroup categories were defined relative to the nationwide percentage of persons having the characteristic: block groups whose percentages were below the national average for the characteristic were classified as low and those whose percentages were above the national average were classified as high. The cases with missing values for a given characteristic were excluded from the response rate calculation for that characteristic.

## 2.2 Results

As indicated in Table 2-1, the weighted overall predicted response rate for the Household Screener is 57.1 percent. The weighted response rates for demographic subgroups indicate the subgroups differ from one another by as much as 8.9 percentage points. The differences among subgroups on weighted response rates were 8.9 percentage points for education, 2.6 percentage points for race, 3.8 percentage points for ethnicity, and 8.7 percentage points for poverty status. The overall predicted response rate for the Household Screener is lower than the projected rate of 70 percent previously presented to OMB in Attachment 22, but it exceeds the worst-case scenario response rate of 39.7 percent.

---

<sup>4</sup> Information from the 5-year (2008 to 2012) rather than the 1-year (2012) 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 2008 to 2012, are the only estimates from ACS that can provide information at the tract level and smaller geographies (see [http://www.census.gov/acs/www/guidance\\_for\\_data\\_users/estimates/](http://www.census.gov/acs/www/guidance_for_data_users/estimates/)).

Table 2-1. PATH Study baseline predicted response rates for the predictor sample, by address characteristics: Household Screener

Characteristic <sup>a</sup>	A Completed (n)	B Interim likely to be completed <sup>b</sup> (n)	C Finalized nonresponse <sup>c</sup> (n)	D Total interim <sup>d</sup> (n)	Unweighted predicted response rate for baseline, based on predictor sample <sup>e</sup> (%)	Weighted predicted response rate for baseline, based on predictor sample <sup>e</sup> (%)
Overall	5,655	404	3,537	1,398	57.2	57.1
Education (% with Bachelor's degree)						
High > 29.1%	2,118	159	1,707	566	51.9	51.9
Low ≤ 29.1%	3,537	245	1,830	832	61.0	60.8
Race (% Black alone or in combination)						
High > 13.7%	1,360	74	991	231	55.5	55.1
Low ≤ 13.7%	4,295	330	2,546	1,167	57.8	57.7
Ethnicity (% Hispanic)						
High > 16.9%	1,473	132	786	421	59.9	59.9
Low ≤ 16.9%	4,182	272	2,751	977	56.3	56.1
Poverty Status						
High > 15.9%	2,159	146	1,032	467	63.0	62.8
Low ≤ 15.9%	3,496	258	2,505	931	54.2	54.1

Note: The projected response rate for baseline is 70 percent.

<sup>a</sup> The characteristics are as sampled. That is, information on the characteristics was collected in the Household Screener. The information used to define the subgroups is from the 5-year (2008 to 2012) American Community Survey.

<sup>b</sup> Interim likely to be completed is the sum of : (1) the product of the number of ever refused interim cases and the estimated proportion of ever refused interim cases that will ultimately result in completes, and (2) the product of the number of never refused interim cases and the estimated proportion of never refused interim cases that will ultimately result in completes.

<sup>c</sup> Finalized nonresponse includes refused cases and all other nonresponding cases.

<sup>d</sup> Total interim includes ever refused interim cases and never refused interim cases.

<sup>e</sup> Predicted response rate = (A+B)/(A+C+D).

# Response Rates Associated with the Adult and Youth Interviews

## 3

### 3.1 Adult Extended Interview

The Adult Extended Interview gathers information from adults (18 years old and older) about tobacco use behaviors, attitudes, knowledge, and health effects, as well as other information including 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).

#### **Method**

The predictor sample includes 1,999 adults selected for the Adult Extended Interview. As of February 26, 2014, all of the Adult Extended Interview cases were finalized.

Table 3-1 provides overall predicted response rates for the Adult Extended Interview and response rates for tobacco use status<sup>5</sup> and demographic subgroups. All response rates are conditional on a completed Household Screener. The response rates were calculated as the product of (1) the Individual or Phase 2 Screener<sup>6</sup> response rate (which uses the same formula as the Household Screener); 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:

$$RRU = (((C+(X * IR)+(Y * IO))/(C+R+NC+O+1(U)+IR+IO))*(CE/(CE+CX)), \text{ where}$$

- IR = number of ever refused interim cases;
- IO = number of never refused interim cases;
- X = probability of IR cases becoming respondents;
- Y = probability of IO cases becoming respondents;

<sup>5</sup> Tobacco use status is as sampled based on information obtained in the Household Screener.

<sup>6</sup> Adults selected on the basis of the Household Screener were asked to complete the Phase 2 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 99.6 percent completed the Adult Extended Interview.

CE = number of Adult Extended Interview completes; and  
CX = number of Adult Extended Interview nonresponses.

For this report, the PATH Study modeled the probabilities of interim cases becoming respondents, X and Y, using the procedure described in Section 2.1. The mean probabilities of the interim cases are 0.07 for adults who ever refused, and 0.57 for the adults who never refused. Again, the predicted response rates assume all pending cases are eligible.

The adult response rates were weighted to compensate for unequal probabilities of selection due to planned oversampling of individuals with certain characteristics. Person-level weights for adults are the product of the Household Screener IPS weights and individual adult IPS weights, which were calculated as the inverse of the selection probabilities for all adults sampled (responding adults and nonresponding adults). (See Section 5.1 for additional information on weighting.)

In addition to the overall row, the table includes rows on tobacco use status, age, sex, race, and ethnicity subgroups. Information from the Household Screener is used to define the demographic characteristics for the responding and nonresponding adults. Some adults had missing values for these characteristics on the Household Screener. Adults with missing information about tobacco use status were sampled using the selection probabilities associated with tobacco users, and are included in the “sampled as user” row of Table 3-1. The cases with missing values for other characteristics were excluded from the response rate calculation for that characteristic.

## **Results**

As indicated in Table 3-1, the weighted overall predicted response rate for the Adult Extended Interview is 75.7 percent. This overall rate is lower than the projected rate of 85 percent, but it exceeds the worst-case scenario response rate of 58.1 percent previously provided to OMB in Attachment 22.

The findings on the weighted response rates for tobacco use status and demographic subgroups indicate the subgroups differ from one another by as much as 8.5 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 1.3 percentage points for tobacco use status, 5.2 percentage points for age, 1.7 percentage points for sex, 8.5 percentage point for race, and 2.6 percentage points for ethnicity.



Table 3-1. PATH Study baseline predicted response rates for the predictor sample, by respondent characteristics: Adult Extended Interview

Characteristic <sup>a</sup>	Phase 2 Screener				Adult Extended Interview		Unweighted predicted response rate for baseline, based on predictor sample <sup>e</sup> (%)	Weighted predicted response rate for baseline, based on predictor sample <sup>e</sup> (%)
	A P2 Screener, completed (n)	B P2 Screener, interim likely to be completed <sup>b</sup> (n)	C P2 Screener, finalized nonresponse <sup>c</sup> (n)	D P2 Screener, total interim <sup>d</sup> (n)	E Adult Extended, completed (n)	F Adult Extended, finalized nonresponse (n)		
Overall	2,820	169	599	475	1,991	8	76.5	75.7
Tobacco use status								
Sampled as user	1,473	88	307	243	1,301	2	77.0	77.2
Sampled as non-user	1,347	82	292	232	690	6	75.7	74.8
Age <sup>f</sup>								
18-24	733	48	145	122	516		78.1	77.8
25-44	967	65	173	171	748	2	78.5	78.4
45-64	764	41	176	132	524	3	74.6	74.1
65+	339	11	91	40	195	3	73.3	72.3
Sex								
Male	1,396	95	304	258	1,040	5	75.8	75.4
Female	1,424	72	288	213	951	3	77.5	76.2
Race <sup>f</sup>								
White only	2,056	125	443	358	1,482	6	76.0	75.4
Black only or in combination with some other race	464	24	77	58	306	1	81.1	80.2
Other	221	13	61	38	164	1	72.6	72.1
Ethnicity <sup>f</sup>								
Hispanic	491	48	69	115	333	4	78.8	76.8
Non-Hispanic	2,327	118	522	352	1,656	4	76.2	75.6

Note: The projected response rate for baseline is 85 percent.

<sup>a</sup> The characteristics are as sampled. That is, information on the characteristics was collected in the Household Screener.

<sup>b</sup> Interim likely to be completed is the sum of: (1) the product of the number of ever refused interim cases and the estimated proportion of ever refused interim cases that will ultimately result in completes, and (2) the product of the number of never refused interim cases and the estimated proportion of never refused interim cases that will ultimately result in completes.

<sup>c</sup> Finalized nonresponse includes refused cases and all other nonresponding cases.

<sup>d</sup> Total interim includes ever refused interim cases and never refused interim cases.

<sup>e</sup> Predicted response rate = ((A+B)/(A+C+D))\*(E/(E+F)).

<sup>f</sup> 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 age, 39 for race, and 2 for ethnicity.

## 3.2 Youth Interview

The Youth Interview gathers information from youth (12 to 17 years old) on similar topics as those in the Adult Extended Interview. Sampled youth are asked about their tobacco use and attitudes about tobacco. In addition, demographic information is collected and youth are asked about environmental factors, family and peer influences, substance use, and mental health. Field interviewers conducted the interviews in person using ACASI.

### **Method**

The predictor sample includes 1,265 youth selected for the Youth Interview. As of February 26, 2014, 1,070 (84.6%) Youth Interview cases were finalized, and 195 (15.4%) cases were still being followed up in field work.

Table 3-2 provides overall predicted response rates for the Youth Interview and responses rates for demographic subgroups. All response rates are conditional on a completed Household Screener. The response rates were calculated using the same formula as that for the Adult Extended Interview. The same probability of never refused interim cases becoming respondents (0.57) and ever refused interim cases becoming respondents (0.07) were used for the Adult Extended Interview and Youth Interview. (See Section 3.1.)

The youth response rates were weighted to compensate for unequal probabilities of selection due to subsampling of youth in households with more than two youths. Person-level weights for youth are the product of the Household Screener IPS weights and individual youth IPS weights, which were calculated as the inverse of the selection probabilities for all youth sampled (responding youth and nonresponding youth). (See Section 5.1 for additional information on weighting.)

In addition to the overall row, the table includes rows on age, sex, race, and ethnicity subgroups. Information from the Household Screener is used to define the demographic characteristics for the responding and nonresponding youth. Because the PATH Study did not collect information on the tobacco use of youth in the Household Screener, information on response rates for that characteristic is unavailable. Youth with missing values for some of the characteristics on the Household Screener were excluded from the response rate calculation for that characteristic.

**Results**

As indicated in Table 3-2, the weighted overall predicted response rate for the Youth Interview is 81.2 percent. This overall rate is higher than the projected rate of 75 percent in Attachment 22. A worst-case scenario response rate was not specified for the Youth Interview.

The findings on the weighted response rates for demographic subgroups indicate the subgroups differ from one another by as much as 6.1 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 4.1 percentage points for age, 1.3 percentage points for sex, 6.1 percentage point for race, and 3.6 percentage points for ethnicity.

**Table 3-2. PATH Study baseline predicted response rates for the predictor sample, by respondent characteristics: Youth Interview**

Characteristic <sup>a</sup>	A Completed (n)	B Interim likely to be completed <sup>b</sup> (n)	C Finalized nonresponse <sup>c</sup> (n)	D Total Interim <sup>d</sup> (n)	Unweighted predicted response rate for baseline, based on predictor sample <sup>e</sup> (%)	Weighted predicted response rate for baseline, based on predictor sample <sup>e</sup> (%)
Overall	964	59	106	195	80.9	81.2
Age <sup>f</sup>						
12-14	481	23	41	85	83.1	83.7
15-17	475	36	65	104	79.3	79.6
Sex						
Male	475	31	56	99	80.3	80.6
Female	489	28	50	96	81.4	81.9
Race <sup>f</sup>						
White only	703	47	85	153	79.7	80.2
Black only or in combination with some other race	154	7	13	21	85.9	85.7
Other	73	4	7	17	79.4	79.6
Ethnicity						
Hispanic	269	23	42	40	83.3	83.9
Non-Hispanic	695	35	63	155	79.9	80.3

Note: The projected response rate for baseline is 75 percent.

<sup>a</sup> The characteristics are as sampled. That is, information on the characteristics was collected in the Household Screener.

<sup>b</sup> Interim likely to be completed is the sum of: (1) the product of the number of ever refused interim cases and the estimated proportion of ever refused interim cases that will ultimately result in completes, and (2) the product of the number of never refused interim cases and the estimated proportion of never refused interim cases that will ultimately result in completes.

<sup>c</sup> Finalized nonresponse includes refused cases and all other nonresponding cases.

<sup>d</sup> Total interim includes ever refused interim cases and never refused interim cases.

<sup>e</sup> Predicted response rate = (A+B)/(A+C+D).

<sup>f</sup> 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 age and 34 for race.

## Response Rates Associated with the Biospecimen Collections

# 4

This section is on the method and response rates for the collection of buccal cell, urine, and blood samples from adults who completed Adult Extended Interviews. Biospecimens are intended to provide a basis for 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 buccal cell and urine samples; on separate visits, phlebotomists collected the blood samples.

### 4.1 Method

As of February 26, 2014, 1,991 adults in the predictor sample had completed the Adult Extended Interview and were eligible to provide biospecimens. Table 4-1 provides overall predicted unweighted and weighted response rates for the biospecimen collections, and responses rates for tobacco use status and demographic subgroups. All 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 formula was used to compute the projected biospecimen response rates presented in Attachment 22 for the baseline wave. The denominator for the rate, the 1,991 adults who completed the Adult Extended Interview, is the same for each 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.

**Table 4-1. PATH Study baseline predicted response rates for the predictor sample, by respondent characteristics: Biospecimen collections**

Characteristic <sup>a</sup>	A Adult Extended Interviews completed (n)	Biospecimen								
		Buccal			Urine			Blood		
		B Collected (n)	Unweighted predicted response rate for baseline, based on predictor sample <sup>c</sup> (%)	Weighted predicted response rate for baseline, based on predictor sample <sup>c</sup> (%)	B Collected (n)	Unweighted predicted response rate for baseline, based on predictor sample <sup>c</sup> (%)	Weighted predicted response rate for baseline, based on predictor sample <sup>c</sup> (%)	B Collected (n)	Unweighted predicted response rate for baseline, based on predictor sample <sup>c</sup> (%)	Weighted predicted response rate for baseline, based on predictor sample <sup>c</sup> (%)
Overall	1,991	1,408	70.7	69.0	1,253	62.9	61.8	734	36.9	36.9
Tobacco Status										
Sampled as user	1,416	1,039	73.4	72.7	919	64.9	64.5	535	37.8	38.6
Sampled as non-user	575	369	64.2	63.9	334	58.1	58.2	199	34.6	34.6
Age <sup>b</sup>										
18-24	523	389	74.4	74.6	340	65.0	65.4	173	33.1	32.9
25-44	742	535	72.1	71.2	473	63.7	63.0	279	37.6	36.6
45-64	530	352	66.4	64.1	319	60.2	59.3	201	37.9	37.3
65+	195	132	67.7	67.4	121	62.1	60.6	81	41.5	41.6
Sex <sup>b</sup>										
Male	1,034	697	67.4	65.8	625	60.4	59.1	350	33.8	34.2
Female	955	711	74.5	72.3	628	65.8	64.8	384	40.2	39.8
Race <sup>b</sup>										
White only	1,433	1,004	70.1	68.7	894	62.4	61.6	539	37.6	38.1
Black only or in combination with some other race	309	225	72.8	70.1	195	63.1	60.5	116	37.5	35.9
Other	196	143	73.0	69.6	130	66.3	64.9	58	29.6	28.8
Ethnicity <sup>b</sup>										
Hispanic	341	248	72.7	72.6	229	67.2	67.6	121	35.5	36.3
Non-Hispanic	1,616	1,134	70.2	68.1	1,000	61.9	60.5	599	37.1	37.0

Note: Table covers respondents who completed the Adult Extended Interview. The projected response rates for buccal and urine for baseline were 80 percent in Attachment 22; the projected response rate for blood for baseline was 65 percent in Attachment 22.

<sup>a</sup> The characteristics are as reported in the Adult Extended Interview.

<sup>b</sup> The sum of counts for this category do not sum to the overall total due to missing values. The number of missing cases is 1 for age, 2 for sex, 53 for race, and 34 for ethnicity.

<sup>c</sup> Predicted response rate = B/A.

## 4.2 Results

### ***Buccal Cells***

The weighted predicted response rate for buccal cells is 69.0 percent, the projected response rate is 80 percent, and the worst-case response rate is 73 percent. The differential weighted response rate for subgroups of respondents ranges from 1.4 percentage points for ethnicity to 10.5 percentage points for age. The response rate for buccal cell collection based on the predictor sample is lower than projected and than the worst-case scenario discussed in Attachment 22.

### ***Urine***

The weighted predicted response rate for urine is 61.8 percent, the projected response rate is 80 percent, and the worst-case response rate is 49 percent. The differential weighted response rate for subgroups of respondents ranges from 4.4 percentage points for race to 6.3 percentage points for tobacco use status. The response rate for urine collection based on the predictor sample is lower than projected, but it exceeds the worst-case scenario discussed in Attachment 22.

### ***Blood***

The weighted predicted response rate for blood is 36.9 percent, the projected response rate is 65 percent, and the worst-case response rate is 39 percent. The differential weighted response rate for subgroups of respondents ranges from 4.0 percentage points for tobacco use status to 9.3 percentage points for race. The response rate for blood collection based on the predictor sample is lower than projected and than the worst-case scenario in Attachment 22.

## Nonresponse Bias Analysis

This nonresponse bias analysis investigates possible differences between estimates calculated from 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.

### 5.1 Method

Section 1.2 describes the selection of the predictor sample that is used as the basis of the nonresponse bias analysis in this section. The predictor sample consists of 455 segments with representation from all 156 PSUs in the PATH Study sample.

Assessment of potential nonresponse bias begins by comparing estimates of demographic counts from the predictor sample with corresponding estimates from the American Community Survey (ACS). The 1-year (2012) ACS estimates, calculated from the 2012 ACS Public Use Microdata Sample (PUMS), were used for comparison purposes. These estimated demographic counts from the ACS PUMS excluded institutional group quarters and persons in noninstitutional group quarters who are not college students. These exclusions correspond 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. Responses to the PATH Study questions on current cigarette smoking, however, 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, 2012 (NHIS); the National Survey on Drug Use and Health, 2012 (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 and question ordering among the surveys.

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 were calculated for all households sampled (responding households and nonresponding households) as the inverse of the selection probability:

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

where  $P_{ijk}$  is the probability that household  $k$  in segment  $j$  of PSU  $i$  is selected to be in the sample. For the predictor 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 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 were calculated as

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



The sampling of adults is performed in two phases. Phase 1 selects adults based on responses to the Household Screener. The probability that adult  $l$  in the household is selected for the Phase 1 sample is 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 household respondent. Adults sampled at Phase 1 are individually asked questions about their age, race, and tobacco usage, and are subsampled for Phase 2 on the basis of their responses to these questions. The adults subsampled for Phase 2 are then administered the Adult Extended Interview. The probability in the formula for AIPSWT is the product of the first-phase and second-phase selection probabilities.

Note that no nonresponse adjustments are performed for the calculation of IPS weights. The weights HHIPSWT, YIPSWT, and AIPSWT are used for all calculations employing IPS weights that are reported in Section 5.2. For the tables presented in Section 5.2, the unweighted counts include categories for missing values. The 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.

Rao-Scott tests for goodness of fit (Rao and Scott, 1981, 1984, 1987) are used to assess the statistical significance of differences between demographic estimates from the PATH Study predictor sample and the comparison quantities from the 2012 ACS (using the 1-year estimates, as described at the beginning of this section). The test assumes that the quantities calculated from the ACS are fixed values without sampling error; therefore, the p-values for the Rao-Scott tests reported in this document use slightly underestimated standard errors. This means that the p-values may be slightly smaller than they would be if sampling errors in the ACS were taken into account. Small p-values indicate that the estimates from the PATH Study predictor sample are significantly different from the quantities in the ACS. Different criteria may be used for what is considered “small” and in general, the interpretation of a p-value from a goodness of fit test depends on the sample size (see for example Royall, 1986); the p-value for an effect size of three percentage points will be smaller for a sample of 10,000 than for a sample of size 100. In this report, p-values less than 0.05 are considered to indicate significant differences. A p-value greater than 0.05 indicates no significant difference between the PATH Study estimate and the comparison quantity from the ACS, and, thus, no reason to conclude that bias due to that characteristic would affect the PATH Study.

Confidence intervals are provided for estimates of cigarette smoking prevalence from the predictor sample. These are constructed using the weight AIPSWT for adult estimates and weight YIPSWT for youth estimates. The PATH Study estimates may then be compared to other surveys by determining whether the point estimate from the external survey falls within the 95 percent

confidence interval constructed from the PATH Study.<sup>7</sup> Taylor linearization is used to calculate the variance, incorporating the complex sampling features of stratification and clustering. SAS software version 9.3 (SAS Institute, 2011) was used to calculate all estimates.

## 5.2 Results

The first set of tables looks at estimates derived from the Household Screener. The demographic quantities are estimated using the roster of household members, with their characteristics provided by the household respondent. The household-level IPS weight HHIPSWT is used in Tables 5-1 through 5-5 to evaluate potential nonresponse bias. If nonresponse is not associated with demographic characteristics, then the percentages calculated using HHIPSWT will be close to those from the ACS.

Table 5-1 presents the unweighted counts and estimated population percentages of adults in the four race/age domains used for sampling adults within households. These counts are from the enumeration of adults done in the Household Screener. The ACS provides comparison quantities for these four domains. The IPS-weighted estimates of percentages in each of the four domains, calculated excluding the missing values, are similar to the 1-year 2012 ACS estimates. The Rao-Scott p-value for goodness of fit is 0.82, indicating that the PATH Study estimates are not significantly different from the ACS percentages. No evidence was found to indicate nonresponse bias with respect to these four demographic domains.

---

<sup>7</sup> This method will give a guideline for the correspondence between PATH Study estimates of smoking and those obtained from other surveys. It does not include the sampling error from the external surveys, however, and therefore does not exactly correspond to a significance test comparing the two surveys. A confidence interval for the difference between the PATH Study estimate and the estimate from another survey would be wider than the confidence intervals reported here, because it would also account for the sampling error of the other survey.

**Table 5-1. Race by age distribution, based on the household enumeration**

Race and age classification	Unweighted count	Weighted percentage, using household IPS weights	Percentage from ACS PUMS
Black* 18-24	233	2.1%	2.1%
Black* 25+	1,238	10.9%	10.3%
Non-Black 18-24	1,269	11.2%	10.9%
Non-Black 25+	8,401	75.8%	76.7%
Missing age or race	332		
Total	11,473	100.0%	100.0%
p-value		0.82	

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

Table 5-2 compares the sex of the adults enumerated on the PATH Study household rosters with the 1-year 2012 ACS distribution. The Rao-Scott p-value for goodness of fit is 0.66, indicating that the PATH Study estimates are not significantly different from the ACS percentages.

**Table 5-2. Distribution of male and female adults listed in the household enumeration**

Sex	Unweighted count	Weighted percentage for adults, using household IPS weights	Percentage from ACS PUMS
Male	5,478	47.8%	48.0%
Female	5,970	52.2%	52.0%
Missing	25		
Total	11,473		
p-value		0.66	

Table 5-3 compares the distribution of household size for the responding households with the independent estimates of those quantities derived from the 1-year 2012 ACS. The PATH Study appears to be obtaining fewer single-person households than occur in the ACS (p-value < 0.0001). The PATH Study also has a lower percentage of single-adult households (Table 5-4) and, probably related to this pattern, a slightly higher percentage of households with youth ages 12-17 than found in the ACS (Table 5-5). Surveys commonly achieve a slightly lower percentage of one-person households because they have fewer members available for contact.<sup>8</sup> 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 6.

<sup>8</sup> See Brault (2013), who found a similar pattern in the CPS ASEC content test. Data collection for the predictor sample of the PATH Study is not finalized, so this distribution may change as more data are collected.

Table 5-3. Distribution of household size based on households responding to the Household Screener

Number of persons in household who are not on active duty	Unweighted count	Weighted percentage, using household IPS weights	Percentage from ACS PUMS
0-1*	1,308	23.2%	27.9%
2	1,803	31.8%	33.7%
3	1,005	17.8%	15.7%
4	851	15.0%	13.0%
5+	688	12.3%	9.7%
Total	5,655	100.0%	100.0%
p-value		< 0.0001	

\*A small number of households contain only emancipated youth and/or adults on active duty, and hence contribute to the zero part of this category.

Table 5-4. Distribution of number of adults based on households responding to the Household Screener

Number of adults in household who are not on active duty	Unweighted count	Weighted percentage, using household IPS weights	Percentage from ACS PUMS
0-1	1,620	28.8%	33.8%
2	2,860	50.8%	50.6%
3+	1,156	20.4%	15.6%
Missing	19		
Total	5,655	100.0%	100.0%
p-value		< 0.0001	

Table 5-5. Distribution of number of youth ages 12-17 based on households responding to the Household Screener

Number of youth ages 12-17 in household	Unweighted count	Weighted percentage, using household IPS weights	Percentage from ACS PUMS
0	4,679	82.9%	84.3%
1	651	11.6%	11.2%
2+	306	5.4%	4.5%
Missing	19		
Total	5,655	100.0%	100.0%
p-value		0.01	

Tables 5-6 and 5-7 are based on adults in the predictor sample responding to the Adult Extended Interview. The PATH Study oversamples young adults, African-American adults, and 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 5-6 presents the estimated race, ethnicity, and sex/age distributions from adults in the predictor sample responding to the Adult Extended Interview. Additional columns in the table present the weighted distributions, using weight AIPSWT, for the adults from whom urine, buccal, and/or blood specimens were collected.

The IPS-weighted estimates of percent male/female are close to the 1-year 2012 ACS percentages, for the adults in the predictor sample responding to the Adult Extended Interview and for those providing each type of biological specimen. Persons ages 25-44 are overrepresented among the adults responding to the Adult Extended Interview, however, and among those who provide urine and buccal cell specimens. The nonresponse-adjusted weights in Section 6, which calibrate to age groups, correct for this discrepancy.

Table 5-6 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 not significantly different from the 1-year 2012 ACS estimates of those quantities. The race distributions of adults who provide urine and buccal cells also accord with the ACS distribution. Hispanic adults, however, are significantly more likely to provide urine or buccal cell specimens.

Table 5-7 compares Adult Extended Interview respondents and those from whom biological specimens were collected on other quantities that are measured in the ACS: education level and presence of health insurance. The adults responding to the Adult Extended Interview, and those contributing biological specimens, are approximately equally likely to have health insurance as respondents to the 2012 ACS. The education level of the adults responding to the Adult Extended Interview, however, tends to be higher than that in the ACS, although that is not the case for the adults contributing biospecimens. In general, education level is associated with tobacco use status (Agaku et al., 2014); the nonresponse-adjusted weights described in Section 6 adjust for educational attainment.

**Table 5-6. Demographic distributions based on adults responding to the Adult Extended Interview, and on adults from whom urine, buccal, and/or blood specimens were collected**

	Adult respondents to Adult Extended Interview		Adults from whom urine specimen is collected		Adults from whom buccal specimen is collected		Adults from whom blood specimen is collected		Percentage from ACS PUMS
	Unweighted count	Weighted percentage, using adult IPS weights	Unweighted count	Weighted percentage, using adult IPS weights	Unweighted count	Weighted percentage, using adult IPS weights	Unweighted count	Weighted percentage, using adult IPS weights	
<b>Sex</b>									
Male	1,034	48.7%	641	47.0%	716	46.7%	357	45.5%	48.0%
Female	955	51.3%	642	53.0%	730	53.3%	398	54.5%	52.0%
Missing	2		0		0		0		
Total	1,991		1,283		1,446		755		100.0%
p-value		0.65		0.60		0.50		0.33	
<b>Age group</b>									
18-24	523	12.2%	352	13.0%	405	13.5%	182	11.7%	13.0%
25-44	742	40.0%	482	40.7%	547	41.1%	285	37.7%	34.4%
45-64	530	31.6%	328	30.6%	362	29.5%	206	32.4%	34.8%
65+	195	16.2%	121	15.7%	132	16.0%	82	18.3%	17.8%
Missing	1		0		0		0		
Total	1,991		1,283		1,446		755		
p-value		0.001		0.002		0.0004		0.44	
<b>Race</b>									
Black, alone or in combination	309	13.7%	195	12.7%	225	13.2%	116	12.5%	12.4%
White alone	1,433	76.2%	917	76.5%	1,033	76.6%	557	79.8%	76.0%
Other	196	10.1%	135	10.8%	149	10.2%	60	7.7%	11.6%
Missing	53		36		39		22		
Total	1,991		1,283		1,446		755		
p-value		0.40		0.86		0.60		0.07	
<b>Ethnicity</b>									
Hispanic	341	17.3%	241	19.8%	264	19.2%	127	17.4%	14.7%
Non-Hispanic	1,616	82.7%	1,018	80.2%	1,156	80.8%	614	82.6%	85.3%
Missing	34		24		26		14		
Total	1,991		1,283		1,446		755		
p-value		0.13		0.01		0.02		0.27	

**Table 5-7. Comparison of education level and health insurance status based on adults responding to the Adult Extended Interview, and on adults from whom urine, buccal, and/or blood specimens were collected**

	Adult respondents to Adult Extended Interview		Adults from whom urine specimen collected		Adults from whom buccal specimen collected		Adults from whom blood specimen collected		Percentage from ACS PUMS
	Unweighted count	Weighted percentage, using adult IPS weights	Unweighted count	Weighted percentage, using adult IPS weights	Unweighted count	Weighted percentage, using adult IPS weights	Unweighted count	Weighted percentage, using adult IPS weights	
<b>Education</b>									
< HS	258	11.9%	192	14.3%	203	13.2%	129	15.4%	13.4%
HS or GED	550	24.3%	342	22.9%	407	24.6%	206	25.0%	28.0%
Some college, no degree	716	33.5%	470	34.6%	533	34.8%	263	33.0%	31.6%
Bachelor degree	299	18.3%	185	18.2%	200	17.3%	100	16.5%	17.3%
> Bachelor degree	151	12.1%	89	9.9%	100	10.2%	57	10.0%	9.7%
Missing	17		5		3		0		0.0%
Total	1,991		1,283		1,446		755		100.0%
p-value		0.02		0.10		0.27		0.62	
<b>Health insurance</b>									
Yes	1,532	83.2%	988	82.6%	1,113	82.7%	590	83.2%	82.9%
No	438	16.8%	288	17.4%	326	17.3%	163	16.8%	17.1%
Missing	21		7		7		2		0.0%
Total	1,991		1,283		1,446		755		100.0%
p-value		0.79		0.81		0.91		0.87	

Table 5-8 presents the estimates of prevalence of current cigarette smoking<sup>9</sup> for the adults responding to 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 present external estimates of smoking prevalence from TUS-CPS, NHIS, NHANES, and NSDUH, respectively, along with 95 percent confidence intervals from those surveys. The estimates of smoking prevalence from each survey were calculated excluding 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, and mode of administration.

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 audio computer-assisted self-interviewing (ACASI) administration for the tobacco usage questions so that the interviewer does not see the 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, NSDUH is a cross-sectional survey on substance use (including tobacco use) and health, including mental health, and the PATH Study is a longitudinal cohort study of tobacco use behaviors and health. Other differences among the questions used in the instruments of these different studies are outlined in Appendix A.

---

<sup>9</sup> 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 given in Appendix A.



Table 5-8. 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]	Percentage from 2010-2011 TUS-CPS [95% confidence interval]	Percentage from 2012 NHIS [95% confidence interval]	Percentage from 2011-2012 NHANES* [95% confidence interval]	Percentage from 2012 NSDUH, original definition** [95% confidence interval]	Percentage from 2012 NSDUH, modified definition [95% confidence interval]
Current smoker	1,989	35.8%	18.5% [16.5%, 20.4%]	16.1% [15.8%, 16.3%]	18.0% [17.4%, 18.6%]	19.8% [17.5%, 22.1%]	23.8% [23.1%, 24.5%]	21.9% [21.2%, 22.7%]
Current smoker, male	1,033	36.2%	19.7% [17.4%, 21.9%]	18.0% [17.7%, 18.4%]	20.4% [19.5%, 21.3%]	23.9% [20.7%, 27.1%]	26.7% [25.7%, 27.7%]	24.9% [23.7%, 26.0%]
Current smoker, female	954	35.5%	17.4% [14.9%, 19.8%]	14.2% [13.9%, 14.5%]	15.8% [15.0%, 16.5%]	16.0% [13.5%, 18.5%]	21.1% [20.1%, 22.1%]	19.3% [18.5%, 20.1%]
Current smoker, age 18-24	522	28.2%	21.2% [18.1%, 24.4%]	17.1% [16.4%, 17.8%]	17.3% [15.4%, 19.1%]	20.4%*** [13.7%, 27.1%]	NA****	NA
Current smoker, age 25-44	742	42.2%	22.0% [18.7%, 25.2%]	17.9% [17.5%, 18.4%]	21.5% [20.4%, 22.6%]	23.3% [20.0%, 26.7%]	NA	NA
Current smoker, age 45-64	529	39.5%	18.6% [15.7%, 21.5%]	17.8% [17.4%, 18.2%]	19.5% [18.5%, 20.5%]	21.3% [18.3%, 24.2%]	NA	NA
Current smoker, age 65+	195	22.6%	7.6% [5.1%, 10.1%]	7.8% [7.5%, 8.2%]	8.9% [8.0%, 9.7%]	9.2% [6.7%, 11.7%]	NA	NA
Current smoker, Hispanic	341	26.1%	12.7% [9.8%, 15.5%]	10.9% [10.4%, 11.5%]	12.5% [11.3%, 13.7%]	16.6% [13.7%, 19.5%]	18.6% [17.0%, 20.2%]	15.5% [14.1%, 17.0%]
Current smoker, white non-Hispanic	1,197	38.7%	18.8% [16.2%, 21.4%]	17.5% [17.2%, 17.8%]	19.6% [18.9%, 20.4%]	20.2% [17.0%, 23.3%]	25.1% [24.2%, 26.0%]	23.9% [23.0%, 24.8%]
Current smoker, other non-Hispanic	415	34.9%	22.1% [17.8%, 26.5%]	NA [NA]	16.7% [15.6%, 17.7%]	20.8% [16.6%, 24.9%]	22.8% [21.1%, 24.6%]	20.2% [18.6%, 21.9%]
Current smoker, every day	1,989	28.3%	14.7% [13.1%, 16.2%]	12.7% [12.4%, 12.9%]	14.1% [13.6%, 14.6%]	16.4% [14.3%, 18.4%]	NA	NA
Current smoker, some days	1,989	7.5%	3.8% [3.1%, 4.5%]	3.4% [3.3%, 3.5%]	3.9% [3.6%, 4.2%]	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, 2013); 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 2012.

Table 5-8 indicates the IPS-weighted estimates of current smoking from the PATH Study are most similar to estimates from NHIS and NHANES. The value from at least one of these surveys is inside each of the 95 percent confidence intervals constructed from the PATH Study estimates.<sup>10</sup> The estimates from TUS-CPS tend to be below the estimates from the PATH Study, NHIS, and NHANES; the 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 estimates fall well within the range of estimates from comparable surveys.

Table 5-9 gives estimates of current cigarette smoking for the adults from whom urine, buccal, and/or blood specimens were collected. The IPS-weighted estimates of smoking are slightly higher for adults who contribute one of the biospecimens, but the differences are not statistically significant. The confidence intervals for smoking among adults providing biospecimens are in line with the estimates from external surveys. This pattern will continue to be monitored, and if needed, an extra step of weighting for nonresponse may be performed for the analysis of biological specimens, as described in Section 6.1.

Results in Tables 5-6 through 5-9 are based on adults in the predictor sample responding to the Adult Extended Interview. Similar analyses were performed for the youth respondents. The demographic estimates are given in Table 5-10 and estimates of cigarette smoking are given in Table 5-11.

Table 5-10 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 2012 ACS percentages. The PATH Study estimate of the percent of youth who are Hispanic, however, is approximately 7 percentage points higher than the corresponding estimate from ACS, indicating that Hispanic youth are more likely to respond to the PATH Study survey.

---

<sup>10</sup> If a 95% 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 5-9. 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]	Percentage from 2010-2011 TUS-CPS [95% confidence interval]	Percentage from 2012 NHIS [95% confidence interval]	Percentage from 2011-2012 NHANES [95% confidence interval]	Percentage from 2012 NSDUH, original definition* [95% confidence interval]	Percentage from 2012 NSDUH, modified definition [95% confidence interval]
Adult respondent to Adult Extended Interview	1,989	18.5% [16.5%, 20.4%]	16.1% [15.8%, 16.3%]	18.1% [17.4%, 18.6%]	19.8% [17.5%, 22.1%]	23.8% [23.1%, 24.5%]	21.9% [21.2%, 22.7%]
Adults providing urine	1,281	20.7% [18.3%, 23.1%]	16.1% [15.8%, 16.3%]	18.1% [17.4%, 18.6%]	19.8% [17.5%, 22.1%]	23.8% [23.1%, 24.5%]	21.9% [21.2%, 22.7%]
Adults providing buccal	1,445	20.9% [18.5%, 23.2%]	16.1% [15.8%, 16.3%]	18.1% [17.4%, 18.6%]	19.8% [17.5%, 22.1%]	23.8% [23.1%, 24.5%]	21.9% [21.2%, 22.7%]
Adults providing blood	755	21.5% [18.3%, 24.6%]	16.1% [15.8%, 16.3%]	18.1% [17.4%, 18.6%]	19.8% [17.5%, 22.1%]	23.8% [23.1%, 24.5%]	21.9% [21.2%, 22.7%]

\* 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 5-10. Demographic distributions based on youth ages 12-17 who completed the Youth Interview

	Unweighted count	Weighted percentage, using youth IPS weights	Percentage from ACS PUMS
<b>Sex</b>			
Male	475	49.0%	51.0%
Female	489	51.0%	49.0%
Missing	0		
Total	964	100.0%	100.0%
p-value		0.25	
<b>Age group</b>			
12-13	340	35.2%	33.7%
14-17	624	64.8%	66.3%
Missing	0		
Total	964	100.0%	100.0%
p-value		0.28	
<b>Race/ethnicity</b>			
Hispanic	278	29.3%	21.9%
Non-Hispanic white alone	482	51.4%	55.2%
Non-Hispanic other	184	19.3%	22.9%
Missing	20		
Total	964	100.0%	100.0%
p-value		0.0002	

Table 5-11. 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]	Percentage from 2011-2012 NHANES [95% confidence interval]	Percentage from 2012 NSDUH [95% confidence interval]	Percentage from 2012 NYTS [95% confidence interval]
Ever tried cigarette smoking, even one or two puffs	964	14.7%	14.7% [12.2%, 17.2%]	20.5% [17.5%, 23.6%]	17.4% [16.7%, 18.1%]	25.6% [23.6%, 27.6%]
Ever tried smoking, male	475	14.5%	14.4% [11.3%, 17.5%]	21.1% [15.9%, 26.3%]	18.4% [17.4%, 19.4%]	27.2% [25.0%, 29.3%]
Ever tried smoking, female	489	14.9%	15.0% [11.6%, 18.4%]	20.0% [14.6%, 25.5%]	16.4% [15.5%, 17.3%]	24.0% [21.8%, 26.2%]
Ever tried smoking, age 12-13	340	5.3%	5.0% [2.9%, 7.1%]	5.6% [1.9%, 9.4%]	4.8% [4.2%, 5.4%]	11.8% [10.2%, 13.4%]
Ever tried smoking, age 14-17	624	19.9%	20.0% [16.5%, 23.4%]	28.3% [23.5%, 33.0%]	23.5% [22.5%, 24.7%]	32.5% [30.0%, 34.9%]
Have smoked in past 30 days	962	3.5%	3.4% [2.4%, 4.4%]	6.9% [4.0%, 9.8%]	6.6% [6.2%, 7.0%]	8.7% [7.7%, 9.8%]

\* Defined as ever tried a cigarette, even one or two puffs.

Table 5-11 estimates one common measure of cigarette smoking prevalence among youth respondents, along with 95 percent confidence intervals. These are compared with estimates from NHANES, NSDUH, and NYTS.<sup>11</sup> Different measures of smoking are used in this report for youth than for adults. The measure of cigarette smoking used for youth is whether the youth has ever tried smoking a cigarette, even one or two puffs (see Appendix A).

Differences 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 usage 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 much lower for a telephone survey of youth smoking than in 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).

Based on the predictor sample, the PATH Study's estimates of the youth smoking measure appear to be slightly lower than the 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), cigarette smoking among teens is dropping (from 2011 to 2012, it dropped by 0.8 percentage points among 12-13 year olds, 1.3 percentage points among 14-15 year olds, and 2.5 percentage points among 16-17 year olds). 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.

---

<sup>11</sup> TUS-CPS does not interview persons younger than 18 about tobacco use.

# Statistical Approach for Addressing Nonresponse

## 6

### 6.1 Computation of Nonresponse-Adjusted Weights

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 quantities such as age, race, and sex to values calculated from the 1-year 2012 ACS (which are considered to be highly accurate because of the large sample size and high response rate for the ACS). These adjustments correct for disparities among these demographic quantities and also for other disparities that might be associated with the demographic quantities. 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***

The household IPS weights were computed for all sampled addresses in the predictor sample. However, some sampled addresses cannot be located/accessed, others are found to be ineligible (e.g., vacant lots and group quarters), and some eligible households do not complete the Household Screener. Adjustments were therefore made to the IPS weights of responding households to compensate for the estimated number of nonresponding households that were eligible for the PATH Study based on all the addresses in the sample for which eligibility status was determined. This eligibility adjustment was done 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,<sup>12</sup> and the percentage of population that is Hispanic in the census block containing the address. Census region and the urbanicity of the PSU were also used when forming the weighting classes.

---

<sup>12</sup> Black is defined as Black alone, or in combination with other races.

Then, within a weighting class, the IPS weights for the responding households were inflated proportionately so that they produce the same sum as the sum of the IPS weights of the responding and nonresponding households combined. 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 then raked to the 1-year 2012 ACS household counts by census region, tenure, and number of persons in the household. For raking purposes, tenure and the number of persons were imputed for households missing this information using logical or hot-deck imputation.<sup>13</sup> The final raked household weight is

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

### **Person Nonresponse-Adjusted Weights**

The raked household-level weight is 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 differs for adults and youth, as described in Section 5.1.

Similarly to the adjustment for Household Screener nonresponse, a nonresponse adjustment was performed to account for nonrespondents to the Adult Extended Interview. The weights of respondents to the Adult Extended Interview were inflated to account for the nonrespondents.

---

<sup>13</sup> See Lohr (2010) for a brief description of raking and imputation methods.



For youth, the initial weights (YBWT) were raked to population totals from the 1-year 2012 ACS, using Census region, age, race/ethnicity, and sex as raking variables. These variables were imputed, either from the Household Screener or using hot-deck imputation, if they were missing. After raking, the final weights for youth are denoted as YRKWT.

The final weights for adults were computed in three steps. First, a nonresponse adjustment was performed using the tobacco use status, age, and sex reported in the Household Screener, separately within the four Census regions. The resulting adult weight, adjusted for nonresponse between Phases 1 and 2 of the adult sampling procedure, for respondents to the Phase 2 Screener, 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 find 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 2012 ACS. The raking was done using combinations of Census region, age, race/ethnicity, sex, and educational attainment. The final raked weight is

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

The adult raked weight ARKWT is also used for the analysis of adults in the predictor sample who provide biospecimens. An additional stage of weighting may be used for biospecimens for the full sample if needed, in which the weights are adjusted to accord with the adults responding to the Adult Extended Interview and then re-raked to the independent population totals from the 1-year 2012 ACS. This adjustment would be performed separately for each type of biospecimen.

This section described the weighting procedure used for the predictor sample. A similar, though not identical, weighting procedure will be used for the full sample from the PATH Study.

## 6.2 Results

In this section, results are presented on the evaluation of the performance of the nonresponse-adjusted weights for variables of interest in the PATH Study. Tables 6-1 through 6-11 repeat the analyses used to produce Tables 5-1 through 5-11, this time using the nonresponse-adjusted weights described in Section 6.1. The estimates calculated using IPS weights are retained in these tables to facilitate easy comparison of the estimates obtained using the two sets of weights. A p-value is given for each of the IPS-weighted and raked-weighted estimates in each of Tables 6-1 through 6-7. The p-value reported for the IPS-weighted estimate is the same as that given in the corresponding table in Section 5: it assesses the statistical significance of the difference between the IPS-weighted estimate from the PATH Study and the 1-year 2012 ACS quantity. The p-value reported for the weighted percentage using the raked weights is for the comparison of the raked-weighted estimate from the PATH Study to the same ACS quantity.

**Table 6-1. Race by age distribution, based on household enumeration**

<b>Race and age classification</b>	<b>Unweighted count</b>	<b>Weighted percentage, using household IPS weights</b>	<b>Weighted percentage, using household raked weights</b>	<b>Percentage from ACS PUMS</b>
<b>Black* 18-24</b>	<b>233</b>	<b>2.1%</b>	<b>1.7%</b>	<b>2.1%</b>
<b>Black* 25+</b>	<b>1,238</b>	<b>10.9%</b>	<b>10.5%</b>	<b>10.3%</b>
<b>Non-Black 18-24</b>	<b>1,269</b>	<b>11.2%</b>	<b>9.4%</b>	<b>10.9%</b>
<b>Non-Black 25+</b>	<b>8,401</b>	<b>75.8%</b>	<b>78.3%</b>	<b>76.7%</b>
<b>Missing age or race</b>	<b>332</b>			
<b>Total</b>	<b>11,473</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>
<b>p-value</b>		<b>0.82</b>	<b>0.09</b>	

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

The household raked weight HHRKWT adjusts the weights so that they agree with the 1-year 2012 ACS household counts by region, tenure, and household size. They would therefore not be expected to bring person-level percentages of specific demographic groups closer to the ACS values. Tables 6-1 and 6-2 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. These estimated percentages are not significantly different from the ACS quantities, although the IPS-weighted percentages are closer to the ACS values. Table 6-6 demonstrates the effect of the person-level weighting adjustments: using the raked adult weights, the race/age and sex distributions are practically identical to those from the ACS.

Tables 6-3 through 6-5 examine the estimates of household size using the raked household weights; as expected, the raked weights bring the estimated percentages in line with the 1-year 2012 ACS values.

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

Sex	Unweighted count	Weighted percentage for adults, using household IPS weights	Weighted percentage for adults, using household raked weights	Percentage from ACS PUMS
Male	5,478	47.8%	47.4%	48.0%
Female	5,970	52.2%	52.6%	52.0%
Missing	25			
Total	11,473	100.0%	100.0%	100.0%
p-value		0.66	0.11	

**Table 6-3. Distribution of household size based on households responding to the Household Screener**

Number of persons in household who are not on active duty	Unweighted count	Weighted percentage, using household IPS weights	Weighted percentage, using household raked weights	Percentage from ACS PUMS
0-1*	1,308	23.2%	28.4%	27.9%
2	1,803	31.8%	33.5%	33.7%
3	1,005	17.8%	15.5%	15.7%
4	851	15.0%	12.9%	13.0%
5+	688	12.3%	9.7%	9.7%
Total	5,655	100.0%	100.0%	100.0%
p-value		< 0.0001	0.98	

\*A small number of households contain only emancipated youth and/or adults on active duty, and hence contribute to the zero part of this category.

**Table 6-4. Distribution of number of adults based on households responding to the Household Screener**

Number of adults in household who are not on active duty	Unweighted count	Weighted percentage, using household IPS weights	Weighted percentage, using household raked weights	Percentage from ACS PUMS
0-1	1,620	28.8%	34.5%	33.8%
2	2,860	50.8%	50.6%	50.6%
3+	1,156	20.4%	14.9%	15.6%
Missing	19			
Total	5,655	100.0%	100.0%	100.0%
p-value		< 0.0001	0.50	

**Table 6-5. Distribution of number of youth ages 12-17 based on households responding to the Household Screener**

<b>Number of youth ages 12-17 in household</b>	<b>Unweighted count</b>	<b>Weighted percentage, using household IPS weights</b>	<b>Weighted percentage, using household raked weights</b>	<b>Percentage from ACS PUMS</b>
0	4,679	82.9%	84.7%	84.3%
1	651	11.6%	10.6%	11.2%
2+	306	5.4%	4.8%	4.5%
Missing	19			
<b>Total</b>	<b>5,655</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>
<b>p-value</b>		<b>0.01</b>	<b>0.31</b>	

Tables 6-6 and 6-7 present the estimates of demographic characteristics, education, and health insurance based on adult respondents to the predictor sample, using the adult raked weight ARKWT. The raking corrects for the slight overestimate in percentage of the 25-44 age group when the IPS weights are used. Notably, 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 age distribution in line with the 1-year 2012 ACS figures for the adults who provide each type of biospecimen as well. The raked estimates of percentages in each race group accord with the ACS percentages for adults completing the extended interview and for adults providing buccal cell or urine specimens. The race distribution for the adults providing blood specimens, however, is marginally significantly different from the ACS values when the raked weights are used. The PATH Study will continue to monitor this for the full sample and if this pattern persists, may construct an additional set of weights for analyzing the blood collection data. Table 6-7 shows that the raked estimates are not significantly different from the ACS distributions for education and health insurance. For the adults responding to the extended interview, the IPS-weighted estimates for percentages of adults at different education levels are significantly different from the ACS comparison quantities, as indicated by the p-value of 0.02. The p-value for comparing the raked-weighted estimates for education with the ACS estimates is 0.72, indicating that the raked weights correct for the disparity in education.

Estimates of smoking prevalence in Table 6-8 using the raked weight ARKWT are very similar to the estimates using the IPS weight AIPSWT, and both of these are in the range of values obtained by other surveys. Table 6-9 gives estimates, using both sets of weights, of current cigarette smoking prevalence for the adults from whom urine, buccal, and/or blood specimens were collected. No additional weighting adjustments were performed to account for nonresponse to the biospecimen

collections. The raked estimates, using the subsets of respondents who provide each type of biospecimen, are similar to the IPS-weighted estimates but are slightly closer to the estimated smoking prevalence that is calculated using all adults responding to the Adult Extended Interview.

Tables 6-10 and 6-11 examine the effect of the raked weight YRKWT on estimates calculated from the responding youth. The raked weights correct for the slight overrepresentation of Hispanics among the youth in the predictor sample. They have little effect, however on the other demographic characteristics (for which the IPS-weighted estimates already agreed with the 1-year 2012 ACS figures) and estimates of smoking prevalence.

**Table 6-6. Demographic distributions based on adults responding to the Adult Extended Interview, and on adults from whom urine, buccal, and/or blood specimens were collected**

	Adult respondents to Adult Extended Interview			Adults from whom urine specimen is collected			Adults from whom buccal specimen is collected			Adults from whom blood specimen is collected			Percentage from ACS PUMS
	Unweighted count	Weighted percentage using adult IPS weights	Weighted percentage using adult raked weights	Unweighted count	Weighted percentage using adult IPS weights	Weighted percentage using adult raked weights	Unweighted count	Weighted percentage using adult IPS weights	Weighted percentage using adult raked weights	Unweighted count	Weighted percentage using adult IPS weights	Weighted percentage using adult raked weights	
<b>Sex</b>													
Male	1,034	48.7%	48.1%	641	47.0%	46.8%	716	46.7%	46.0%	357	45.5%	44.6%	48.0%
Female	955	51.3%	51.9%	642	53.0%	53.2%	730	53.3%	54.0%	398	54.5%	55.4%	52.0%
Missing	2			0			0			0			
Total	1,991			1,283			1,446			755			100.0%
p-value		0.65	0.97		0.60	0.54		0.50	0.27		0.33	0.13	
<b>Age group</b>													
18-24	523	12.2%	13.0%	352	13.0%	13.7%	405	13.5%	14.4%	182	11.7%	12.4%	13.0%
25-44	742	40.0%	34.5%	482	40.7%	35.6%	547	41.1%	35.9%	285	37.7%	32.5%	34.4%
45-64	530	31.6%	34.7%	328	30.6%	33.6%	362	29.5%	32.4%	206	32.4%	35.7%	34.8%
65+	195	16.2%	17.8%	121	15.7%	17.1%	132	16.0%	17.3%	82	18.3%	19.5%	17.8%
Missing	1			0			0			0			
Total	1,991			1,283			1,446			755			
p-value		0.001	0.99		0.002	0.81		0.0004	0.44		0.44	0.73	
<b>Race</b>													
Black, alone or in combination	309	13.7%	13.0%	195	12.7%	12.2%	225	13.2%	12.6%	116	12.5%	11.7%	12.4%
White alone	1,433	76.2%	77.4%	917	76.5%	77.0%	1,033	76.6%	76.9%	557	79.8%	81.0%	76.0%
Other	196	10.1%	9.6%	135	10.8%	10.8%	149	10.2%	10.4%	60	7.7%	7.3%	11.6%
Missing	53			36			39			22			
Total	1,991			1,283			1,446			755			
p-value		0.40	0.38		0.86	0.90		0.60	0.77		0.07	0.05	
<b>Ethnicity</b>													
Hispanic	341	17.3%	15.0%	241	19.8%	17.6%	264	19.2%	17.1%	127	17.4%	15.9%	14.7%
Non-Hispanic	1,616	82.7%	85.0%	1,018	80.2%	82.4%	1,156	80.8%	82.9%	614	82.6%	84.1%	85.3%
Missing	34			24			26			14			
Total	1,991			1,283			1,446			755			
p-value		0.13	0.84		0.01	0.14		0.02	0.21		0.27	0.63	

**Table 6-7. Comparison of education level and health insurance status based on adults responding to the Adult Extended Interview, and on adults from whom urine, buccal, and/or blood specimens were collected**

	Adult respondents to Adult Extended Interview			Adults from whom urine specimen collected			Adults from whom buccal specimen collected			Adults from whom blood specimen collected			Percentage from ACS PUMS
	Unweighted count	Weighted percentage using adult IPS weights	Weighted percentage using adult raked weights	Unweighted count	Weighted percentage using adult IPS weights	Weighted percentage using adult raked weights	Unweighted count	Weighted percentage using adult IPS weights	Weighted percentage using adult raked weights	Unweighted count	Weighted percentage using adult IPS weights	Weighted percentage using adult raked weights	
<b>Education</b>													
< HS	258	11.9%	13.0%	192	14.3%	15.6%	203	13.2%	14.3%	129	15.4%	16.4%	13.4%
HS or GED	550	24.3%	28.1%	342	22.9%	26.3%	407	24.6%	28.0%	206	25.0%	28.3%	28.0%
Some college, no degree	716	33.5%	31.4%	470	34.6%	32.7%	533	34.8%	32.8%	263	33.0%	31.1%	31.6%
Bachelor degree	299	18.3%	16.4%	185	18.2%	16.4%	200	17.3%	15.7%	100	16.5%	14.8%	17.3%
> Bachelor degree	151	12.1%	11.1%	89	9.9%	8.9%	100	10.2%	9.3%	57	10.0%	9.4%	9.7%
Missing	17			5			3			0			0.0%
Total	1,991			1,283			1,446			755			100.0%
p-value		0.02	0.72		0.10	0.52		0.27	0.78		0.62	0.48	
<b>Health insurance</b>													
Yes	1,532	83.2%	83.9%	988	82.6%	82.9%	1,113	82.7%	83.0%	590	83.2%	84.2%	82.9%
No	438	16.8%	16.1%	288	17.4%	17.1%	326	17.3%	17.0%	163	16.8%	15.8%	17.1%
Missing	21			7			7			2			0.0%
Total	1,991			1,283			1,446			755			100.0%
p-value		0.79	0.44		0.81	0.99		0.91	0.95		0.87	0.51	

Table 6-8. 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 2012 NHIS [95% confidence interval]	Percentage from 2011-2012 NHANES* [95% confidence interval]	Percentage from 2012 NSDUH, original definition** [95% confidence interval]	Percentage from 2012 NSDUH, modified definition [95% confidence interval]
Current smoker	1,989	35.8%	18.5% [16.5%, 20.4%]	18.1% [16.0%, 21.1%]	16.1% [15.8%, 16.3%]	18.0% [17.4%, 18.6%]	19.8% [17.5%, 22.1%]	23.8% [23.1%, 24.5%]	21.9% [21.2%, 22.7%]
Current smoker, male	1,033	36.2%	19.7% [17.4%, 21.9%]	20.0% [17.4%, 22.6%]	18.0% [17.7%, 18.4%]	20.4% [19.5%, 21.3%]	23.9% [20.7%, 27.1%]	26.7% [25.7%, 27.7%]	24.9% [23.7%, 26.0%]
Current smoker, female	954	35.5%	17.4% [14.9%, 19.8%]	16.3% [13.9%, 18.7%]	14.2% [13.9%, 14.5%]	15.8% [15.0%, 16.5%]	16.0% [13.5%, 18.5%]	21.1% [20.1%, 22.1%]	19.3% [18.5%, 20.1%]
Current smoker, age 18-24	522	28.2%	21.2% [18.1%, 24.4%]	21.3% [17.9%, 24.7%]	17.1% [16.4%, 17.8%]	17.3% [15.4%, 19.1%]	20.4%*** [13.7%, 27.1%]	NA****	NA
Current smoker, age 25-44	742	42.2%	22.0% [18.7%, 25.2%]	22.7% [19.2%, 26.1%]	17.9% [17.5%, 18.4%]	21.5% [20.4%, 22.6%]	23.3% [20.0%, 26.7%]	NA	NA
Current smoker, age 45-64	529	39.5%	18.6% [15.7%, 21.5%]	17.3% [14.3%, 20.2%]	17.8% [17.4%, 18.2%]	19.5% [18.5%, 20.5%]	21.3% [18.3%, 24.2%]	NA	NA
Current smoker, age 65+	195	22.6%	7.6% [5.1%, 10.1%]	8.5% [5.7%, 11.4%]	7.8% [7.5%, 8.2%]	8.9% [8.0%, 9.7%]	9.2% [6.7%, 11.7%]	NA	NA
Current smoker, Hispanic	341	26.1%	12.7% [9.8%, 15.5%]	13.0% [9.9%, 16.2%]	10.9% [10.4%, 11.5%]	12.5% [11.3%, 13.7%]	16.6% [13.7%, 19.5%]	18.6% [17.0%, 20.2%]	15.5% [14.1%, 17.0%]
Current smoker, white non-Hispanic	1,197	38.7%	18.8% [16.2%, 21.4%]	18.2% [15.6%, 20.8%]	17.5% [17.2%, 17.8%]	19.6% [18.9%, 20.4%]	20.2% [17.0%, 23.3%]	25.1% [24.2%, 26.0%]	23.9% [23.0%, 24.8%]
Current smoker, other non-Hispanic	415	34.9%	22.1% [17.8%, 26.5%]	21.6% [17.7%, 25.5%]	NA	16.7% [15.6%, 17.7%]	20.8% [16.6%, 24.9%]	22.8% [21.1%, 24.6%]	20.2% [18.6%, 21.9%]
Current smoker, every day	1,989	28.3%	14.7% [13.1%, 16.2%]	14.6% [12.9%, 16.4%]	12.7% [12.4%, 12.9%]	14.1% [13.6%, 14.6%]	16.4% [14.3%, 18.4%]	NA	NA
Current smoker, some days	1,989	7.5%	3.8% [3.1%, 4.5%]	3.5% [2.8%, 4.1%]	3.4% [3.3%, 3.5%]	3.9% [3.6%, 4.2%]	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, 2013); 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 2012.



Table 6-9. 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 2012 NHIS [95% confidence Interval]	Percentage from 2011-2012 NHANES [95% confidence Interval]	Percentage from 2012 NSDUH, original definition* [95% confidence Interval]	Percentage from 2012 NSDUH, modified definition [95% confidence Interval]
Adult respondent to Adult Extended Interview	1,989	18.5% [16.5%, 20.4%]	18.1% [16.0%, 20.1%]	16.1% [15.8%, 16.3%]	18.1% [17.4%, 18.6%]	19.8% [17.5%, 22.1%]	23.8% [23.1%, 24.5%]	21.9% [21.2%, 22.7%]
Adults providing urine	1,281	20.7% [18.3%, 23.1%]	19.9% [17.4%, 22.4%]	16.1% [15.8%, 16.3%]	18.1% [17.4%, 18.6%]	19.8% [17.5%, 22.1%]	23.8% [23.1%, 24.5%]	21.9% [21.2%, 22.7%]
Adults providing buccal	1,445	20.9% [18.5%, 23.2%]	20.3% [17.8%, 22.8%]	16.1% [15.8%, 16.3%]	18.1% [17.4%, 18.6%]	19.8% [17.5%, 22.1%]	23.8% [23.1%, 24.5%]	21.9% [21.2%, 22.7%]
Adults providing blood	755	21.5% [18.3%, 24.6%]	19.9% [16.8%, 23.1%]	16.1% [15.8%, 16.3%]	18.1% [17.4%, 18.6%]	19.8% [17.5%, 22.1%]	23.8% [23.1%, 24.5%]	21.9% [21.2%, 22.7%]

\*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 6-10. Demographic distributions based on youth ages 12-17 who completed the Youth Interview

	Unweighted count	Weighted percentage, using youth IPS weights	Weighted percentage, using youth raked weights	Percentage from ACS PUMS
<b>Sex</b>				
Male	475	49.0%	51.0%	51.0%
Female	489	51.0%	49.0%	49.0%
Missing	0			
Total	964	100.0%	100.0%	100.0%
p-value		0.25	0.99	
<b>Age group</b>				
12-13	340	35.2%	35.6%	33.7%
14-17	624	64.8%	64.4%	66.3%
Missing	0			
Total	964	100.0%	100.0%	100.0%
p-value		0.28	0.20	
<b>Race/ethnicity</b>				
Hispanic	278	29.3%	22.3%	21.9%
Non-Hispanic white alone	482	51.4%	54.9%	55.2%
Non-Hispanic other	184	19.3%	22.7%	22.9%
Missing	20			
Total	964	100.0%	100.0%	100.0%
p-value		0.0002	0.97	

Table 6-11. 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 2012 NSDUH [95% confidence interval]	Percentage from 2012 NYTS [95% confidence interval]
Ever tried cigarette smoking, even one or two puffs	964	14.7%	14.7% [12.2%, 17.2%]	14.7% [12.0%, 17.5%]	20.5% [17.5%, 23.6%]	17.4% [16.7%, 18.1%]	25.6% [23.6%, 27.6%]
Ever tried smoking, male	475	14.5%	14.4% [11.3%, 17.5%]	14.7% [11.2%, 18.1%]	21.1% [15.9%, 26.3%]	18.4% [17.4%, 19.4%]	27.2% [25.0%, 29.3%]
Ever tried smoking, female	489	14.9%	15.0% [11.6%, 18.4%]	14.8% [11.0%, 18.5%]	20.0% [14.6%, 25.5%]	16.4% [15.5%, 17.3%]	24.0% [21.8%, 26.2%]
Ever tried smoking, age 12-13	340	5.3%	5.0% [2.9%, 7.1%]	5.2% [2.6%, 7.8%]	5.6% [1.9%, 9.4%]	4.8% [4.2%, 5.4%]	11.8% [10.2%, 13.4%]
Ever tried smoking, age 14-17	624	19.9%	20.0% [16.5%, 23.4%]	20.0% [16.2%, 23.7%]	28.3% [23.5%, 33.0%]	23.5% [22.5%, 24.7%]	32.5% [30.0%, 34.9%]
Have smoked in past 30 days	962	3.5%	3.4% [2.4%, 4.4%]	3.5% [2.3%, 4.7%]	6.9% [4.0%, 9.8%]	6.6% [6.2%, 7.0%]	8.7% [7.7%, 9.8%]

\* Defined as ever tried a cigarette, even one or two puffs.

This report by NIDA/FDA addresses the terms of clearance of OMB's approval (0925-0664 dated August 23, 2013) of the PATH Study's baseline wave of data and biospecimen collection. It covers the first 5 months of the baseline (September 12, 2013 to February 26, 2014) and is based on the predictor sample, the probability sample of addresses selected for the PATH Study and released to field interviewers early in the field period.

## 7.1 Summary of Findings

### **Response Rates**

As reported in Sections 2, 3, and 4, the weighted response rates<sup>14</sup> for two of the PATH Study interviews and the biospecimen collections based on the predictor sample are lower than projected (see Table 6-1), and the weighted response rates for two biospecimen collections are slightly below the worst-case scenario rates for the full sample provided in Attachment 22.

**Table 7-1. Summary of PATH Study baseline overall response rates for the predictor sample**

Collection	Unweighted predicted response rate, based on predictor sample	Weighted predicted response rate, based on predictor sample	Projected response rate*	Worst-case scenario response rate*
Household Screener	57.2%	57.1%	70%	39.7%
Adult Extended Interview	76.5%	75.7%	85%	58.1%
Youth interview	81.0%	81.2%	75%	–
Buccal cell	70.7%	69.0%	80%	73%
Urine	62.9%	61.8%	80%	49%
Blood	36.9%	36.9%	65%	39%

\*Provided in the request to OMB for baseline data and biospecimen collection.

The differential weighted response rates for tobacco use status and demographic subgroups are generally modest. (See Tables 2-1, 3-1, 3-2, and 4-1.) The largest differential weighted response rate, 10.5 percentage points, is for age for buccal cell collection; this differential rate suggests a

<sup>14</sup> These response rates were weighted with inverse probability of selection weights.

heightened potential for nonresponse bias. Notably, the differential weighted response rates for blood collection, which range from 4.0 percentage points for tobacco use status to 9.3 percentage points for race, are consistent with those for the other collections.

As discussed in Section 1.2, the PATH Study based the interim report on a predictor sample designed to estimate results for the entire baseline sample. Although this approach ensures a large proportion of the cases were finalized by the time the report analyses were conducted, it does not fully reflect the important improvements to the Study implemented months after the predictor sample was fielded. Those changes, which are intended to boost response rates, include enhanced field interviewer training on obtaining biospecimen consent, improved coordination of blood collection visits, and extensive efforts to identify and schedule field work for times potential respondents are most likely to be available. In addition, the substantial experience gained by the field interviewers with the predictor sample and other early sample releases is expected to have increased their effectiveness with later sample releases. For these reasons, the response rates for the predictor sample are likely to underestimate those for the entire baseline sample.

### ***Nonresponse Bias Analysis***

Nonresponse bias analysis shows that estimates of most of the key demographic and tobacco use variables calculated from the PATH Study predictor sample with the inverse probability of selection weights are comparable to those produced by other national general population and health surveys. However, the completed interviews from the predictor sample to date appear to underrepresent single-person households relative to the 1-year 2012 ACS counts.

Based on the predictor sample, estimated percentages of demographic characteristics for adults completing the Adult Extended Interview and for adults contributing biospecimens are not significantly different from the 1-year 2012 ACS values for most characteristics. The estimated percentages of adults who are Hispanic are similar to ACS values for adults responding to the Adult Extended Interview and for adults who provide blood specimens, but Hispanics are overrepresented among adults who provide urine and buccal cell specimens. In addition, the estimated percentage of adults who are 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 provide urine and buccal cell specimens. Adults responding to the Adult Extended Interview in the predictor sample also exhibit somewhat higher education levels than in the ACS. These differences are not apparent among adults who provide biospecimens, however.

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 predictor sample are roughly mid-way in the range of estimates on smoking. Hence, the analyses found no evidence of nonresponse bias with respect to this important measure.

Estimates of demographic characteristics of youth from the predictor sample agree with 1-year 2012 ACS values for most demographic characteristics. However, Hispanic youth are overrepresented among PATH Study respondents. (The nonresponse weight adjustments correct for this overrepresentation.)

PATH Study estimates of the selected youth cigarette smoking measure from the predictor 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 these surveys are from 2011 and 2012 while those from the PATH Study are from the first 5 months of the baseline wave, September 12, 2013 to February 26, 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 5.2, time period is one of a number of factors that may explain the different estimates. Estimates of cigarette smoking prevalence among youth have large confidence intervals for all of the surveys studied.

### ***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 predictor sample indicate the nonresponse adjustments are successful for reducing the discrepancy between the PATH Study estimates and 1-year estimates from the 2012 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.

Estimates of adult cigarette smoking and health insurance coverage 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 correct for the slight overrepresentation of Hispanics among youth in the predictor sample but have little effect

on the other demographic characteristics (i.e., IPS-weighted estimates already agreed with the ACS values) and estimates of youth cigarette smoking.

## 7.2 Conclusions and Implications for Study Going Forward

### **Conclusions**

NIDA concludes that the PATH Study baseline wave of data and biospecimen collection is yielding scientifically defensible results that will meet study objectives. The response rates for two of the three data collections, Household Screener and Adult Extended Interview, are lower than projected. However, nonresponse bias analysis found the characteristics of the respondents are generally in line with the 1-year estimates from the 2012 ACS. Estimates of the cigarette smoking rate among adults based on the predictor sample are within the range of rates found in other national health studies. Moreover, when the predictor sample estimates were adjusted for nonresponse using the raked weights, they more closely approximated the ACS estimates and the adult smoking rates remained essentially the same.

The response rate for the third data collection, Youth Interview, is higher than projected. For this collection, the nonresponse bias analysis also found the characteristics of respondents to be generally consistent with the 1-year estimates from the 2012 ACS. The ever-tried-smoking rate for youth based on the predictor sample is at the low end of the range of rates found by other national health studies. However, when the predictor sample estimates were adjusted for nonresponse, they more closely approximated the 2012 ACS estimates and the ever-tried-smoking rates for youth found by other national studies.

The response rates for the three biospecimen collections are lower than projected, and the response rates for the buccal cell and blood collections are slightly below the worst-case scenario rates. Nonetheless, nonresponse bias analysis found the characteristics of the respondents to be generally in line with estimates from the 1-year 2012 ACS. When the predictor sample estimates were adjusted for nonresponse, they more closely approximated the ACS estimates.

Due to the limited number of predictor sample biospecimens that have been analyzed to date, this report does not include a comparison of predictor sample biospecimen results (e.g., on nicotine metabolites) with those from other national studies that collect survey data on smoking in

combination with biospecimens. Analyses of biospecimens are continuing, however, with plans for these and other analyses in the near future for the predictor sample biospecimens and eventually, for the full baseline sample.

### ***Implications for Study Going Forward***

The implications of these findings for the PATH Study are that ongoing efforts to increase the response rates should be explored and implemented; adjustments to the sampling strategy and/or target yields should be considered to compensate for the lower response rates achieved to date; and the approach for adjusting the IPS weights to account for nonresponse should be continued and refined for the full sample. Each of these three courses of action is further discussed below.

First, the PATH Study will seek to increase its response rates during the baseline wave. As mentioned in Section 7.1, the Study has been implementing steps intended to improve its response rates. Some of these have taken place after the 5<sup>th</sup> month of the baseline, however, and are not fully reflected in the predictor sample results. The PATH Study is continuously seeking ways to improve its response rates in the baseline. In addition, as discussed in Supporting Statement B of the PATH Study's non-substantive change request for the baseline wave, the Study has developed steps to help it achieve high response rates in its followup waves. These involve maintaining contact with baseline respondents, tracing respondents for whom contact is lost, and reaching out to engage individuals who age into the youth cohort or adult cohort.

Second, the PATH Study is planning to adjust its sampling strategy to compensate for the lower response rates achieved to date and any revisions to target baseline sample sizes. This strategy may include releasing additional addresses to the field during the field period and increasing the sampling rates for adults at the household and individual screening phases. As needed, the Study will adjust its analytic plans to account for potentially smaller sample yields than planned, for example, by combining some subgroups.

Third, once the baseline wave has ended, the PATH Study plans to continue its approach to adjusting the IPS weights to account for nonresponse. Doing so for the interim report highlighted the usefulness of this approach in reducing potential nonresponse bias. The Study will continue to refine and improve the weighting procedures as more data become available. The Study will also repeat the nonresponse bias analysis that was conducted for this report. When performed with the full sample, the nonresponse bias analysis will serve the same purposes as in this report: to provide



measures of the Study's validity and contribute to refining the weighting procedures for the full sample.

## 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.
- 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](http://www.copafs.org/UserFiles/file/fcsm/H3_Brault_2013FCSM.pdf).
- Center for Behavioral Health Statistics and Quality. (2013a). 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. (2013b). Results from the 2012 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/NSDUH/2012SummNatFindDetTables/DetTabs/NSDUH-DetTabsCover2012.pdf>.
- Centers for Disease Control and Prevention. (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](http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6245a2.htm).
- 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.
- Gooley, T.A., Leisenring, W., Crowley, J., and Storer, B.E. (1999). Estimation of Failure Probabilities in the Presence of Competing Risks: New Representations of Old Estimators. *Statistics in Medicine*, 18, 695–706.
- Lohr, S. (2010). *Sampling: Design and Analysis*, 2<sup>nd</sup> 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](http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf).
- Rao, J.N.K. and Scott, A.J. (1981). The Analysis of Categorical Data from Complex Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables. *Journal of the American Statistical Association*, 76, 221–230.

- Rao, J.N.K. and Scott, A.J. (1984). On Chi-Squared Tests for Multiway Contingency Tables with Cell Properties Estimated from Survey Data. *The Annals of Statistics*, 12, 46–60.
- Rao, J.N.K. and Scott, A.J. (1987). On Simple Adjustments to Chi-Square Tests with Survey Data. *The Annals of Statistics*, 15, 385–397.
- Royall, R.M. (1986). The effect of sample size on the meaning of significance tests. *The American Statistician*, 40, 313–315.
- 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.
- 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.fcs.gov/13papers/I3\\_Song\\_2013FCSM.pdf](http://www.fcs.gov/13papers/I3_Song_2013FCSM.pdf), and slides available at [http://www.copafs.org/UserFiles/file/fcs.gov/I3\\_Song\\_2013FCSM.pdf](http://www.copafs.org/UserFiles/file/fcs.gov/I3_Song_2013FCSM.pdf).
- Substance Abuse and Mental Health Services Administration (SAMHSA, 2013). *Results from the 2012 National Survey on Drug Use and Health: Summary of National Findings*, NSDUH Series H-46, HHS Publication No. (SMA) 13-4795. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from <http://www.samhsa.gov/data/NSDUH/2012SummNatFindDetTables/NationalFindings/NSDUHresults2012.htm#ch4>, <http://www.samhsa.gov/data/NSDUH/2012SummNatFindDetTables/DetTabs/NSDUH-DetTabsSect2peTabs1to42-2012.htm>.
- United Nations (2005). *Designing Household Survey Samples: Practical Guidelines*. United Nations Publication ST/ESA/STAT/SER.F/98, New York: United Nations. Available at <http://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf>.
- United States Department of Commerce, Census Bureau (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>.
- 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.fcs.gov/05papers/Wang\\_Aldworth\\_etal\\_VIB.pdf](http://www.fcs.gov/05papers/Wang_Aldworth_etal_VIB.pdf).

# 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.

**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)*
<b>Question to define current smoking (answers defining current smoking given in parentheses)</b>					
“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)
<b>Age range included in estimate</b>					
18+	18+	18+	20+	18+	18+
<b>Exclusions from population</b>					
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.	
<b>Proxy responses allowed</b>					
No	Yes	Yes, for individuals physically or mentally incapable of responding (468 cases in 2012)*	No	No	No

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

\*\* Proxies are allowed if 4<sup>th</sup> 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](http://appliedresearch.cancer.gov/studies/tus-cps/surveys/tuscps_english_2010.pdf), p3.

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

PATH Study	NHANES	NSDUH	NYTS
<b>Question to define ever tried cigarette smoking (answers defining ever tried cigarette smoking given in parentheses)</b>			
“Have you ever tried cigarette smoking, even one or two puffs?” (yes)	<p>“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))</p> <p>I have never smoked, not even a puff (1), 1 or more puffs but never a whole cigarette (2),</p> <p>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)</p>	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)
<b>Questions for determining whether have smoked in past 30 days</b>			
“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)
<b>Ages of youth in survey</b>			
12-17	12-17	12-17	12-17 year old students in public or private schools
<b>Exclusions from population</b>			
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.
<b>Other comments</b>			
	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 estimates are given in Center for Behavioral Health Statistics and Quality (2013a, b) gives estimates and the standard errors of the estimates.	The survey is administered by teachers in the classroom setting.