

**Supporting Statement B  
for  
Population Assessment of  
Tobacco and Health (PATH) Study (NIDA) –  
Second Wave of Data Collection**

**June 26, 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>
B.	Collections of Information Employing Statistical Methods.....	1
B.1	Respondent Universe and Sampling Methods.....	1
B.2	Procedures for the Collection of Information.....	7
B.3	Methods to Maximize Response Rates and Deal with Nonresponse.....	25
B.4	Test of Procedures or Methods to be Undertaken.....	30
B.5	Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data.....	33
	References.....	34

### LIST OF ATTACHMENTS (Appearance in Supporting Statement B)

Attachment 9. Follow-up, Retention, and Tracking Materials.....	11
Attachment 19. Field Data Collection Materials.....	11
Attachment 12. Consent Materials.....	12
Attachment 2. PATH Study Data Collection Instruments.....	15
Attachment 21. PATH Study Interim Report.....	26

Attachment 22. List of Statistical Consultants..... 31

## **B. Collections of Information Employing Statistical Methods**

The following section focuses on a description of the statistical methods planned for Wave 2 of the PATH Study. Section B.1 describes the baseline target population of the PATH Study as well as the respondent universe and the expected baseline and Wave 2 sample compositions by various age, tobacco-use, and race-ethnic subgroups. It also includes an overview of the baseline and Wave 2 sample designs. The section ends with a discussion of the PATH Study's expected response rates for Wave 2. Section B.2 describes the procedures for collecting PATH Study data. It presents weighting and estimation procedures, with an elaboration of the degree of precision expected for the analyses of various domains of interest. Section B.3 describes procedures for maximizing the participation and retention of the PATH Study respondents. Section B.4 discusses procedures for evaluating the data collection procedures, including a discussion of nonresponse bias. The final section, Section B.5, presents a list of statistical consultants contributing to the PATH Study.

### **B.1 Respondent Universe and Sampling Methods**

#### **B.1a Target Population**

The baseline target population of the PATH Study is the civilian household population 18 years of age or older in the U.S. (the 50 states and the District of Columbia), and youth ages 12 to 17. College students are sampled through their permanent residence rather than at their dormitory. Active-duty members of the military (Army, Navy, Marines, Air Force, and Coast Guard) are excluded, as are all persons living in institutional and non-institutional group quarters other than college dormitories. Spouses and children of active-duty military living off post in the 50 states and D.C. are covered.

## B.1b Respondent Universe and Estimated Sample Composition

Estimates of the PATH Study youth respondent universe and estimated respondent sample sizes at baseline and Wave 2 are shown in the second row of Table B-1. Under the planned sample design, the estimated number of completed interviews with youth ages 12 to 17 at baseline is approximately 14,050. The baseline estimates in the table are based on data from the 2012 American Community Survey (ACS) and replicate<sup>1</sup> of the PATH Study. After accounting for baseline shadow sample<sup>2</sup> members who have turned 12, youths interviewed at baseline who have become adults, and expected attrition among the remainder of baseline youth cohort, the estimated number of completed interviews with youth ages 12 to 17 at Wave 2 is approximately 12,642.

Estimates of the PATH Study adult respondent universe are shown in Table B-2, which presents the number of persons by age, tobacco usage, and race domains derived from population projections. There are varying definitions of “tobacco user.” Table B-2 provides estimated baseline sample sizes for each of three definitions of interest for the PATH Study. The first, called the “wide net” definition, classifies a person as a tobacco user if he or she has smoked a cigarette, cigar, or pipe, or used smokeless tobacco in the last 30 days; and/or has ever used an e-cigarette, snus, dissolvable tobacco, or smoked tobacco in a hookah. This “wide net” is intended to capture adults who have had experience with tobacco products and who may be at risk of progressing to more frequent use. A “current user” of tobacco is anyone who (1) has smoked at least 100 cigarettes in their lifetime and smokes cigarettes every day or some days, and/or (2) smokes cigars/cigarillos/pipe and/or uses smokeless tobacco every day or some days, and/or (3) uses e-cigarettes, hookah tobacco, snus, and/or dissolvable

---

<sup>1</sup> 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). The interviews from replicate 1, therefore, are a representative probability sample of the civilian non-institutionalized U.S. population.

<sup>2</sup> The “shadow sample” consists of children who are between the ages of 9 and 11 at the household’s baseline interview in Wave 1. These children are not interviewed at baseline, but will be enrolled in the youth cohort in subsequent waves when they attain age 12.

tobacco every day or some days.<sup>3</sup> Finally, a “current or experimental user” of tobacco is either (a) anyone who is a “current user” or (b) anyone who has used any of these tobacco products in the past month.

The respondent universe counts in the second column of Table B-2 were computed by applying the estimated wide-net tobacco use rates from the predictor sample of the PATH Study<sup>4</sup> to adult civilian household population counts from the 2012 ACS for each age/race domain. Under the current sample design, the estimated number of completed adult interviews at baseline is 31,625, including approximately 8,495 young adults (18 to 24 year-olds) and 5,629 Blacks or African Americans (Black/AA).<sup>5</sup> The number of tobacco users is, of course, largest and the number of non-users is smallest under the “wide net” definition whereas the reverse is true under the “current user” definition. At Wave 2, the corresponding estimates are 29,103 completed adult interviews, with 8,180 young adults and 5,163 Blacks or African Americans. These numbers account for both aging of the baseline sample participants and expected attrition. The PATH Study will generate longitudinal data on wide-ranging tobacco use behaviors within the cohort. Pending the availability of these data, the sample sizes presented in Table B-2 for Wave 2 are estimated using the wide-net tobacco use rates calculated from replicate 1 of the baseline sample.

---

<sup>3</sup> The definition of tobacco use in the Current Population Survey-Tobacco Use Supplement (CPS-TUS) encompasses items (1) and (2) of the “current user” definition, but not item (3). Note that, although the “current user” definition is considered to be scientifically more appropriate for most of the analyses of the PATH Study data, analysts wishing to employ the CPS-TUS definition in their analyses will have the data available to do so.

<sup>4</sup> The predictor sample was a 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. Inverse-probability-of-selection weights and nonresponse-adjusted weights were constructed for the predictor sample, which allowed it to be used for estimating population percentages of tobacco users. The estimates of wide-net tobacco use were calculated using the nonresponse-adjusted weights, and have large standard errors because of the small sample sizes.

<sup>5</sup> Questions in the PATH Study’s instruments that collect data on race or ethnicity will be consistent with the most recent revision of the OMB Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting. However, the term “Black/AA” as used here refers to anyone who chooses African American or Black as a race category (irrespective of whether one or more race categories are chosen and irrespective of their reported ethnicity).

Except for the number of youth in the shadow sample, i.e., 9-to-11 year-olds selected at baseline for the purpose of replenishing the 12-to-17 year-old youth sample in later waves but not for the purpose of interviews, the sample size estimates in Tables B-1 and B-2 apply to the baseline and Wave 2 *completed interviews* (with or without biological specimens for adults). Specific subgroups in these tables represent the major sampling strata used at the person level at baseline. Power projections are provided later in Supporting Statement B for subgroups of potential analytic interest.

Table B-1. PATH Study youth and shadow youth respondent universes and estimated sample sizes at baseline and Wave 2

<b>Group</b>	<b>Respondent universe</b>	<b>Estimated baseline sample size</b>	<b>Estimated Wave 2 sample size</b>
Children 9-11 (shadow sample)	12,639,240	7,158	4,772
Youth 12-17	25,611,322	14,050	12,642

Table B-2. PATH Study adult respondent universes and estimated sample sizes at baseline and Wave 2

Group	Respondent universe under the “wide net” definition of tobacco use	Estimated baseline sample size under the “wide net” definition	Estimated baseline sample size under the “current user” definition	Estimated baseline sample size under the “current or experimental” user definition	Estimated Wave 2 sample size under “wide net” definition
18-24 Black/AA user	2,327,249	1,233	785	951	1,031
18-24 Black/AA non-user	2,624,345	443	891	725	529
18-24 non-Black/AA user	14,945,357	5,334	3,556	3,991	4,728
18-24 non-Black/AA non-user	9,160,057	1,485	3,263	2,827	1,891
25+ Black/AA user	10,722,529	2,692	2,006	2,400	2,465
25+ Black/AA non-user	14,807,303	1,261	1,946	1,553	1,138
25+ non-Black/AA user	57,496,336	13,511	10,116	11,112	12,267
25+ non-Black/AA non-user	122,179,715	5,666	9,061	8,065	5,053
All adults	234,262,891	31,625	31,625	31,625	29,103

### B.1c Sample Design

The baseline sample for the PATH Study was selected using a four-stage, stratified probability sample design involving the selection of: (1) 156 primary sampling units (PSUs) consisting of counties or groups of contiguous counties; (2) 6,049 second-stage sampling units (referred to as segments); (3) 168,857 mailing addresses; and (4) 45,675 eligible persons within households occupying dwelling units (DUs) at sampled addresses.<sup>6</sup> In addition to the four stages of selection, a two-phase approach was used for the fourth stage of sampling of adults within households. Interviews were attempted with all youth ages 12- to-17 and adults sampled at baseline. In addition, a “shadow sample” of youth ages 9-to-11 was selected for use as a refresher sample for the youth cohort in later waves of the study. The sampling frames and methods used at each stage of selection for the

<sup>6</sup> The number of PSUs and segments in the PATH Study is fixed at the time of writing; however the final numbers of sampled mailing addresses and persons may deviate slightly from the estimates provided.

baseline sample are described in Sections B.1c and B.1d of Supporting Statement B for the baseline wave.

The PATH Study currently plans to follow the cohort for three years over the duration of the contract period. Additional follow-up waves, with sample refreshment, are under consideration pending the availability of funding. For Wave 2 of the PATH Study, there are no plans to a) subsample persons for interview from those who were selected and participated in the interview at baseline, or b) add a sample of new participants. Youth in the shadow sample who are permitted by a parent or guardian to participate in the study and have since reached the age of 12 will be interviewed for the first time at Wave 2. Similarly, 17-year olds in the youth sample at baseline who reach age 18 by Wave 2 will receive a baseline version of the adult instrument and be asked to provide urine and blood samples for the first time.

For Wave 2, the PATH Study is planning to subsample approximately 12,500 adults for urine collection from adults who were selected and participated in biospecimen collection at baseline, with the expectation that 10,000 of these adults, plus aging-in youth, will provide a urine sample at Waves 2 and 3. Selection criteria for this subsample are currently being determined by the PATH Study's Biological Working Group based on priority areas in tobacco regulatory science.

The PATH Study's target population at baseline excluded all active-duty members of the military (Army, Navy, Marines, Air Force, and Coast Guard) and all persons living in group quarters other than college dormitories. Some of the baseline sample members will be active duty at Wave 2 and others will have moved into a group quarters living arrangement. All baseline respondents will be retained as members of the PATH Study cohort for Wave 2 and, whenever feasible, every effort will be given to obtaining interviews (and possibly biospecimens) at Wave 2. This may include, for example, waiting for a study participant to return to the household from a short-term group quarters stay before interviewing him/her for Wave 2.

## B.1d Estimated Response Rates

For the baseline wave, response rates for the screener and extended interviews are expected to be 57 percent for households, 75 percent for sampled adults, and 77 percent for sampled youth. The overall baseline response rate is estimated to be 43 percent for adults and 44 percent for youth (i.e., the product of the expected screener response rate and the expected person-level response rate).<sup>7</sup> For older adults (ages 25 and over at baseline), projected response rates for the extended interviews are 86 percent for Wave 2; 87 percent for Wave 3; and 90 percent for future follow-up waves. For younger adults (ages 18 to 24 at baseline), the corresponding response rate assumptions are 85 percent, 85 percent, and 89 percent. The projected response rates for youth are higher, 90 percent for each of Waves 2 and 3, and 92 percent for future follow-up waves. The attrition rates are estimated based primarily on the Medical Expenditure Panel Surveys (MEPS) and the National Longitudinal Survey of Youth (NLSY).<sup>8</sup> The higher levels of attrition for younger adults take into account their higher likelihood of moving residence. Data from the 2012 3-year ACS show that about 30 percent of 18 to 24 year olds

---

<sup>7</sup> The baseline response rates are calculated using results from predictor sample of the PATH Study baseline data collection.

<sup>8</sup> There are few recent in-person longitudinal surveys in the U.S. that are directly comparable to the PATH Study. MEPS, like the PATH Study, conducts in-person interviews with respondents, and provides the most recent data on retention of adults in a longitudinal study. Kashihara and Ezzati-Rice (2004), adjusting for the fact that MEPS interviews are conducted every six months rather than annually, estimated year-1 retention for the 1999-2000 MEPS at 90% and year-2 retention at 95%. The retention rates for more recent years of MEPS have been in line with these published rates. Conservative values are given for the PATH Study retention rates to account for differences between the PATH Study and MEPS (such as differences in the frequency of visits and in incentive amounts. The current National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) is a cross-sectional survey, but the 2001-2002 NESARC had a follow-up wave in 2004-2005 with a retention rate of 86.7 percent (National Institutes of Health, 2010). The projected retention rates are less than or equal to rates given for other longitudinal surveys (National Research Council, 2014), or for the British Household Panel Survey (Contoyannis et al. 2004; Lynn 2006, Table 67, where the wave 2, wave 3, and wave 4 retention rates are 87 percent, 91 percent, and 96 percent, respectively). The wave 1 retention for youth in the National Longitudinal Survey of Youth 1997 (NLSY, 2014) was 93 percent, with higher rates for subsequent waves.

The retention rates also account for the expected mortality between waves, based on 2011 data in Table 1 of [http://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61\\_06.pdf](http://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_06.pdf) (national vital statistics) and 2011 ACS data. Assumptions differ for adults ages 18 to 24 at baseline (100 percent) and adults ages 25 and over at baseline (99.15 percent).

move residence annually, compared to 13 percent of persons ages 25 and over. When combined with an estimate of the percentage of movers who will relocate outside of the PATH Study’s geographic reach,<sup>9</sup> and potentially higher levels of noncooperation, an additional 1-2 percent attrition is anticipated among younger adults at each wave. Nevertheless, the PATH Study will undertake a series of measures in an effort to minimize attrition and achieve the expected response rates, as described in Section B.3. The overall Wave 2 response rate is projected to be approximately 37 percent for adults and 40 percent for youth (i.e., the product of the baseline response rate and the expected Wave 2 response rate).

For Table B-3, the estimated counts of adults providing biospecimens at Wave 2 are based on several assumptions. It is assumed that youth who completed the youth interview at baseline and age into the adult cohort by Wave 2 will be asked to complete the adult interview and to provide urine and blood samples at Wave 2. Among this group, it is assumed that the response rates will be 69 percent for urine and 45 percent for blood.<sup>10</sup> At this time, the PATH Study is planning to continue blood collection among only new adults (aging-in 18-year olds) in Wave 2 and Wave 3. For the purpose of estimating respondent burden, it is assumed that, by definition, the sample of 12,500 adults asked to provide a urine sample at Wave 2 (see Section B.1c) cannot be a youth at baseline who has aged into the adult cohort at Wave 2. Among this sample, it is assumed that 80 percent of adults will cooperate.

Table B-3. Estimated number of respondents for Wave 2

Sampling unit	Percentage or estimated response rate	Estimated number
Primary sampling unit (PSU)		156
Area segments/CDSF segments		6,049
Households with persons sampled at baseline		56,793
<b>Adult sample (persons ages 18+)</b>		

<sup>9</sup> The PATH Study’s geographic reach consists of areas within 100 miles of a study PSU.

<sup>10</sup> These numbers are based on the percentages of adults ages 18 to 24 with completed interviews who have provided urine or blood specimens, respectively, as of June 16, 2014. The percentage who will provide a urine sample is projected to be slightly higher than the observed baseline rate of 66% due to the lower overall biospecimen burden at Wave 2 (given the absence of buccal cell collection).

Number of youths completing baseline interview		14,050
Number from baseline youth sample eligible for Wave 2 adult interview (one-sixth of youth sample)	17%	2,295
Number from baseline youth sample completing Wave 2 adult interview	85%	1,990
Number of adults completing baseline interview		31,625
Number of adults completing baseline and Wave 2 interviews*	86%	27,113
Number of adults completing Wave 2 interview		29,103
Number of adults providing urine specimen at Wave 2		11,373
Number of adults providing blood specimen at Wave 2		896
<b>Youth sample (persons ages 12-17)</b>		
Number of youth permitted to participate in baseline shadow sample		7,158
Number from baseline shadow sample eligible for Wave 2 interview (one-third of children ages 9 to 11 turn 12)	33%	2,338
Number from baseline shadow sample completing youth interview at Wave 2**	88%	2,105
Number of youth completing baseline interview		14,050
Number from baseline youth sample eligible for Wave 2 interview (five-sixths of youth sample)	83%	11,708
Number of youth completing baseline and Wave 2 interviews	90%	10,537
Number of youth completing Wave 2 interviews		12,642

\* The value of 86 percent for the percentage of adults completing the Wave 2 interview is a weighted average of the assumed response rates for adults ages 18 to 24 at baseline (85 percent) and adults ages 25 and over at baseline (86 percent). The number calculated in the table applies the respective assumed response rates to the baseline sample sizes in the two age groups.

\*\*A slightly lower response rate is assumed for the first interview of youth in the shadow sample because Wave 2 is the first time that active participation is requested for these youth.

## B.2 Procedures for the Collection of Information

This section includes a brief overview and a description of the PATH Study's plans to complete extended interviews with youth and adults, to reduce burden associated with redundant data collection, and to collect biospecimens from adults.

## B.2a Overview

The PATH Study Wave 2 data and biospecimen collection involves three main components. These are: (1) automated ACASI (audio computer-assisted self-interviewing) extended instruments (separate instruments for youth and adults), (2) an automated CAPI (computer-assisted personal interviewing) parent instrument, and (3) collection of biospecimens from adults (urine is collected from a subsample of adults who provided a urine sample at baseline; and urine and blood are collected from "aging-in" adults who enrolled as youth at baseline but have now aged into the adult cohort).<sup>11</sup> Collection of biospecimens is not a requirement for adult participation; however, completion of an extended interview at baseline is required. The components and instruments differ for the four sets of study participants in Wave 2: (1) adult sampled persons (SPs), (2) youth SPs and their parents, (3) youth SPs who age into the adult cohort, and (4) shadow youth who age into the youth cohort and their parents.

The primary responsibility of the PATH Study field interviewer is to obtain complete and accurate information from the sampled persons assigned to them. Meeting this responsibility facilitates proper nonresponse analysis. All field interviewers working on the PATH Study receive extensive training on the procedures for administration of the data collection instruments, including techniques to establish rapport and gain cooperation, to explain the study's importance to the respondent, and to answer respondent questions or address any concerns.

The PATH Study provides training to field interviewers in three forms: home study, in-person, and web-based. All *newly hired field interviewers* participate in a 16-hour home study program designed to introduce trainees to the PATH Study. This program focuses on respondent contact materials and provides an opportunity for field interviewers to practice gaining cooperation and establishing rapport. Newly hired field interviewers without previous interviewer experience receive an additional five hours of home study training in

---

<sup>11</sup> During the baseline wave, buccal cells also were collected from adult participants, but this collection was discontinued before the end of the wave due to funding limitations. There are currently no plans to resume buccal cell collection in Wave 2 or Wave 3.

general interviewing techniques. All newly hired field interviewers participate in a 6-day in-person in-depth training in the entire set of data and biospecimen collection procedures, including: (1) techniques for obtaining consent; (2) conducting the ACASI extended interviews and CAPI parent interview; (3) collecting tobacco product information from adults by scanning UPC codes on products; (4) collecting and shipping urine samples; (5) scheduling blood collections; (6) issuing respondent incentives; and (7) completing administrative procedures, such as data transmission and reporting to a field supervisor. *Experienced field interviewers from the baseline wave* participate in a 16-hour web-based training that focuses on new tasks for the Wave 2 collections as well as on a review of ongoing tasks that continue unchanged from the baseline into Wave 2. In addition, field interviewers receive a field procedures manual that provides detailed reference materials on locating addresses, the interviewing process, questionnaire content, and biospecimen collection. Experienced phlebotomists receive training on PATH Study procedures for visiting the homes of consenting adults to collect blood samples; this training includes the phlebotomist manual on collecting blood as part of the PATH Study.

The PATH Study uses three modes of rigorous quality control procedures to ensure that field interviewers are following specified procedures and protocols and that the data collected are of the highest quality: in-person observation, telephone verification, and review of audio recordings of interviews.

First, field interviewers who successfully complete training but show any area of potential weakness are observed in-person at least one time by a supervisor or home office staff member. Observing field interviewers conducting their job in the field is an effective method of performance monitoring, not only of administration of the interview, but also of adherence to the PATH Study's procedures. In-person observations are typically concentrated in the early weeks of data collection so that problems can be detected as early as possible; this provides an opportunity for prompt corrective feedback to the individual field interviewer to help improve his/her on-the-job performance.

Second, brief telephone verification interviews are conducted with a sample of respondents by PATH Study home office staff to confirm that an interview was administered or attempted as reported by the field interviewer. Quality control

standards for the PATH Study require the verification of at least ten percent of each field interviewer's finalized work to ensure that the interview was conducted according to study procedures. This includes cases finalized as complete as well as those with non-complete dispositions, such as refusal.

Third, as part of quality control, selected items from the CAPI interviews and consent discussions with adults are audio-recorded (with the consent of respondents) using CARI (computer-assisted recorded interviewing) and reviewed to assess interviewer performance. As needed (e.g., when a respondent refuses audio-recording), quality control interviews are conducted by telephone by home office staff; for some non-complete dispositions (e.g., dwelling unit is vacant), an experienced, specially trained field interviewer validates the disposition in person.

Additionally, throughout the field period, supervisors remain in close contact with field interviewers. Scheduled weekly telephone conferences are held in which non-finalized cases assigned to field interviewers are reviewed to determine the best approach for working and finalizing the cases.

Management staff at the home office and at remote locations have access to a supervisor management system, including automated management and production reports that are used to monitor the data collection effort. Field interviewers are required to transmit data on a daily basis; data are transmitted to a secure server at the home office to update the automated management reports. These data are used to produce weekly reports on progress during the past week as well as on potentially suspicious field interviewer behavior (e.g., anomalies in the amount of time between interviews, the scheduling of interviews very early in the morning or late in the evening, or the number of interviews conducted per day).

## **B.2b Extended Interview**

The data collection procedures differ for (1) adult sampled persons (SPs), (2) youth SPs and their parents, (3) youth SPs who age into the adult cohort, and (4)

shadow youth who age into the youth cohort and their parents. Approximately 3 months in advance of the anniversary of the baseline interview (planned household interview date), the Home Office mails a letter to adult SPs and parents of youth SPs that reminds them of the upcoming follow-up interview. Approximately 1 month in advance of the planned household interview date, the field interviewer contacts the adult SP and the parent of a youth SP by telephone to arrange a convenient time for the in-person visit at the SP's home; as appropriate, the field interviewers contact the participants by telephone to confirm the appointment.

### **Adult SPs**

At the in-person visit, the field interviewer: (1) reviews the main elements of the informed consent for interview provided by the SP at baseline;<sup>12</sup> (2) administers the adult extended interview, which includes updating contact information about the adult; (3) as appropriate, reviews the main elements of consent for the biospecimen collection obtained during baseline; (4) collects a urine sample from a subsample of adults who provided a urine sample at baseline; and (5) pays the incentive to the respondent. (The biospecimen collection is discussed further in Section B.2d.) If a sample adult is unavailable or unable to complete the interview at the scheduled time, the field interviewer attempts to schedule an appointment for a return visit or, at a minimum, determine the best time for a return visit.

After reviewing the main elements of consent, the field interviewer provides a brief automated tutorial on using ACASI and launches the automated ACASI extended interview. As required throughout the interview, the field interviewer aids the sample person in providing a response. At the end of the extended interview, the field interviewer updates contact information for that person,

The sample adult who completes the extended interview receives \$35 (the adult extended interview incentive) as a thank you for completing the interview. These

---

<sup>12</sup> The consent documents are framed in terms of the baseline wave. This is appropriate for the two audiences that will be exposed to the documents at Wave 2: (1) baseline respondents who may wish to review the consent document they signed at baseline, and (2) aging-in participants for whom Wave 2 is their baseline.

respondents also receive a thank you letter (Attachment 9). A refusal conversion letter is sent to sample adults who initially decline to participate or are difficult to contact (Attachment 19). An adult respondent may also receive \$5 for updating his/her contact information on up to two occasions during the year, for a total of \$10.

### **Youth SPs**

At the in-person visit, the field interviewer: (1) reviews with the parent the main elements of parent permission for the youth to participate, which was obtained at baseline; (2) reviews with the parent the main elements of consent for the short parent interview, also obtained at baseline; and (3) administers the CAPI parent interview, which includes updating the parent's contact information. If a parent of a sampled youth is unavailable or unable to participate at that time, the field interviewer attempts to schedule an appointment for a return visit or, at a minimum, determine the best time for a return visit. The youth interview is not conducted until parental informed consent has been reviewed. The parent who completes a parent interview for the youth receives \$10 as a thank you for completing the interview.

For a selected youth with parental permission, if the youth is available and has an adequate amount of time to complete the interview, the field interviewer reviews with the youth the main elements of assent for the interview, which was obtained at baseline. The interviewer then begins a brief automated tutorial on using ACASI and launches the automated ACASI extended interview for the youth to complete. As required throughout the interview, the field interviewer is available to aid the sample person in the use of ACASI and responding to the interview. If a sample youth is unavailable or unable to complete the interview at the scheduled time, the field interviewer attempts to schedule an appointment for a return visit or, at a minimum, determine the best time for a return visit.

The youth respondent who completes the extended interview receives \$25 (the youth extended interview incentive) as a thank you for completing the interview. The parents of youth respondents receive a thank you letter (Attachment 9). A refusal conversion letter is sent to the parents of respondents who are difficult to

contact (Attachment 19). A youth respondent may also receive \$5 on up to two occasions when his/her parent updates the youth's contact information during the year, for a total of \$10.

### **Youth SPs Who Age into the Adult Cohort**

At the in-person visit, the field interviewer: (1) obtains informed consent (Attachment 12); (2) administers the adult extended interview, which includes gathering additional contact information about the adult; (3) obtains consent for the biospecimen collection; (4) collects the urine sample; (5) arranges a follow-up appointment for a phlebotomist to collect a blood sample; and (6) pays the incentive to the respondent at the completion of the first home visit. (The biospecimen collection is discussed further in Section B.2d.) If a sample adult is unavailable or unable to complete the interview at that time, the field interviewer attempts to schedule an appointment for a return visit or, at a minimum, determine the best time for a return visit.

After obtaining consent, the field interviewer provides a brief automated tutorial on using ACASI and launches the automated ACASI extended interview. As required throughout the interview, the field interviewer is available to aid the sample person in the use of ACASI and responding to the interview. At the end of the extended interview, the field interviewer gathers additional contact information for that person and asks the respondent to consent to providing biospecimens. (See Section B.2d.)

The sample adult who completes the extended interview receives \$35 (the adult extended interview incentive) as a thank you for completing the interview. These respondents also receive a thank you letter (Attachment 9). A refusal conversion letter is sent to sample adults who initially decline to participate or are difficult to contact (Attachment 19). An adult respondent may receive \$5 for updating his or her contact information on up to two occasions during the year, for a total of \$10.

## **Shadow Youth Who Age into the Youth Cohort**

At the in-person visit, the field interviewer: (1) obtains parent permission for the youth to participate; (2) obtains consent for the short parent interview; and (3) administers the CAPI parent interview, which includes updating contact information about the youth from the parent. If a parent of a sampled youth is unavailable or unable to participate at that time, the field interviewer attempts to schedule an appointment for a return visit or, at a minimum, determine the best time for a return visit. The youth interview is not conducted until parental informed consent has been obtained. The parent who completes a parent interview for the youth receives \$10 as a thank you for completing the interview.

For a selected youth with parental permission, if the youth is available and has an adequate amount of time to complete the interview, the field interviewer obtains youth assent (Attachment 12) and then attempts to complete the automated ACASI extended instrument. If a sample youth is unavailable or unable to complete the interview at that time, the field interviewer attempts to schedule an appointment for a return visit or, at a minimum, determine the best time for a return visit.

After obtaining assent from the selected youth, the field interviewer provides a brief automated tutorial on using ACASI and launches the automated ACASI extended interview. As required throughout the interview, the field interviewer is available to aid the sample person in the use of ACASI and responding to the interview.

The youth respondent who completes the extended interview receives \$25 (the youth extended interview incentive) as a thank you for completing the interview. The parents of youth respondents receive a thank you letter (Attachment 9). A refusal conversion letter also is sent to the parents of respondents who are difficult to contact (Attachment 19). A youth respondent may receive \$5 on up to two occasions when his/her parent updates the youth's contact information during the year, for a total of \$10.

## **B.2c Burden Reduction by Avoiding Redundant Data Collection**

The Wave 2 interviews for adults and youth who completed baseline interviews take full advantage of the information collected at baseline. Stable information such as demographic characteristics (e.g., sex and race) is collected only at baseline. Similarly, information on lifetime use of tobacco products is not asked again for products a respondent reported having used at baseline; instead, this information establishes a baseline for updating information on the use of the products since the baseline interview. This approach, of not repeating questions from the baseline about characteristics or behaviors that are unlikely to have changed at Wave 2, helps to keep respondent burden to an average of 1 hour for the adult interview and one-half hour for the youth interview.

The parent interview collects personal information about the parent of a sampled youth, some general characteristics of the household as a whole, and information about the youth, plus contact information to support reaching the parent and youth for future data collection activities. Because more than one youth may be sampled per household, one parent may be asked to respond to a parent interview in regard to more than one youth. In this instance, the parent is not asked to again provide his or her personal information, the household information, or the contact information after the first instance of the parent interview. Rather, the parent is only asked questions that are uniquely relevant to each sampled youth.

## **B.2d Biospecimens**

Under two circumstances, the field interviewer asks adult SPs who complete an extended interview at Wave 2 to provide biospecimens as part of the PATH Study. First, the field interviewers collect urine samples from a subsample of 10,000 adults who provided urine at baseline. (See Table B-4.) Second, the field interviewers collect urine and arrange for collection of blood samples from youth SPs who age into the adult cohort at Wave 2 and consent to biospecimen collection.

Although completion of the extended interview is required from all respondents who choose to participate in the longitudinal cohort, providing biospecimens is voluntary and not a condition of participation.

Table B-4. Summary of plans for biospecimen collection at Wave 2

Type of respondent	Biospecimen	
	Urine	Blood
Subsample of 10,000 adults who provided a urine sample at baseline	Yes	No
Youth SPs who age into adult cohort at Wave 2	Yes	Yes

### Urine

At Wave 2, the field interviewer requests consent for the collection of urine from 18-year olds who have aged into the adult cohort at Wave 2. Following completion of the interview, the field interviewer collects urine specimens from these consenting adults, and from the subsample of adults who provided urine at baseline and were subsampled to provide urine at Wave 2. The field interviewer provides written and oral instructions to the respondent for collection of the urine specimen, then packs the specimen and ships it to the PATH Study’s biorepository.

The respondent who provides urine receives \$25 as a thank you for participating in the urine sample component of the study.

### Blood

At Wave 2, the field interviewer requests consent to collect blood from adults who aged into the adult cohort at Wave 2. For adults who consent to provide blood, the field interviewer schedules an appointment for a visit by a PATH Study phlebotomist to collect the blood specimen. After the initial home visit by the field

interviewer, the phlebotomist contacts the adult to confirm the appointment for collecting the blood specimen.

Upon visiting the respondent's home, the phlebotomist administers the blood suitability exclusion questions (Attachment 2) for blood collection (CAPI instrument) and requests that respondents answer items about his/her recent use of tobacco products (CASI instrument) (Attachment 2). The phlebotomist then collects the blood specimen, and packs and ships it to the PATH Study biorepository.

The respondent who provides a blood specimen during a second home visit receives \$25 as a thank you for participating in the blood sample component of the study.

## **B.2e Weighting and Estimation Procedures**

Cross-sectional and longitudinal sample weights will be developed for the PATH Study respondents to permit estimation for and inference about the population from which the sample is drawn. The sample weights will be produced to accomplish the following objectives:

- Permit the appropriate development of estimates, taking account of the fact that not all persons in the target population have the same probability of selection;
- Limit the potential for biases arising from differences between cooperating and non-cooperating sample persons and households;
- Use auxiliary data on known population characteristics in such a way as to reduce coverage biases and benchmark the PATH Study's estimates to the corresponding population totals;
- Reduce the variation of the weights and prevent a small number of observations from dominating domain estimates; and
- Facilitate sampling error estimation appropriate to the complex sample design.

The data used in weighting will undergo careful edit, frequency, and consistency checks to prevent errors in the sample weights. The checks will be performed on items to be used in the weighting procedures and will be limited to records that require weights. These checks are important because errors in the weights can affect the PATH Study's estimates.

The process for computing (cross-sectional) baseline weights is described in detail in Section B.2e of Supporting Statement B for the baseline wave. The basic steps include:

1. Creating household base weights that are the inverses of the household selection probabilities;
2. Creating household nonresponse-adjusted weights by inflating the household base weights of responding households to compensate for nonresponding households;
3. Creating person base weights by modifying the household nonresponse-adjusted weights to compensate for unequal selection probabilities of sampled persons;
4. Creating person nonresponse-adjusted weights by inflating the person base weights of responding persons to compensate for nonresponding persons;
5. Creating trimmed weights to reduce any excessive variation in the person nonresponse-adjusted weights;
6. Creating final weights by “raking” the trimmed weights to population control totals to account for undercoverage and other sources of bias that may remain after applying the above steps; and
7. Creating replicate weights using the jackknife method for use in variance estimation.

For the PATH Study, one set of baseline weights will be created for all youth who complete the baseline interview and another set will be created for all adults who complete the baseline interview. Those 9-to-11 year-olds selected as part of the shadow sample will be included in the baseline weighting process, as far as Step 3

above. Their person base weights will serve as the “base weights” for the shadow sample members when they become 12 years old and join the youth cohort.

Other sets of weights may be of interest to facilitate analyses of baseline data for specific subsets of adults who provide one or more of the biospecimen samples at baseline. However, a balance must be maintained between the analytic value of creating weights tailored to specific analyses, the resources required to do so, and the user experience in working with data files and choosing appropriate weights. An alternative approach for handling component nonresponse (here, biospecimen nonresponse) at baseline is to treat the component nonresponse as a set of item nonresponses in a respondent record and use imputation<sup>13</sup> as a means of compensation for the missing data. In this case, the analytic data file for the baseline data collection would comprise all sampled adults who completed the interview, irrespective of whether they provided any of the biospecimens, and all sampled youth who completed the interview.

The previous discussion focused on weighting for respondents in the baseline wave. The next section discusses the development of cross-sectional and longitudinal weights for respondents in both the baseline wave and Wave 2. For both waves, specially developed SAS macros will be used to compute the weights for the PATH Study sample. These macros perform such tasks as cell weighting adjustments for nonresponse, poststratification, raking, generalized regression estimation, creation of replicates for variance estimation, and weighting adjustments (i.e., nonresponse adjustment, poststratification, generalized regression estimation, and raking) of the replicate weights.

### **Development of Cross-sectional Weights for Wave 2 Respondents**

Sampled persons who age into the youth or adult cohort study (i.e., reach age 12 or 18) at Wave 2 will be assigned weights for Wave 2 cross-sectional analyses and possibly for some longitudinal analyses thereafter. The starting weights for

---

<sup>13</sup> De Waal et al. (2013) and van Buuren (2012) summarize various imputation methods that are in common use. The “Autoimpute” software developed by the prime contractor (Judkins et al., 2007) preserves the weights of the observations while performing competitively with Bayesian methods for imputation.

baseline shadow sample members who age into the youth cohort at Wave 2 will be their person base weights computed during the baseline weighting process. The starting weights for all youth and adults interviewed at baseline will be their final raked weight computed during the baseline weighting process.

The second step in creating the Wave 2 cross-sectional weights will be to adjust the starting weights for nonresponse between baseline and Wave 2, taking into account baseline data on tobacco usage where available. The nonresponse-adjusted weights will also be poststratified or raked to Wave 2 (i.e., updated) population control totals from the Census Bureau's Population Estimates Program and/or the American Community Survey (ACS). While an adjustment for attrition is more typically associated with the creation of longitudinal weights, it is deemed important for the Wave 2 cross-sectional weighting process as well due to the lack of reliable population control totals regarding tobacco usage. The poststratification/raking process may be conducted separately for 12 year-olds (and possibly 18 year-olds) each year to help ensure that they are fully represented among the youth (and possibly adult) age group.

### **Development of Longitudinal Weights for Wave 2 Respondents**

As a large longitudinal cohort study, the PATH Study expects some attrition among study participants to occur at Wave 2, whether from loss to follow-up or to refusal to participate. Consequently, some form of statistical compensation will be required for missing data. Those who respond at both waves will constitute the data set for longitudinal analyses. Two alternative approaches can be used for compensating for baseline respondents who do not respond at the second wave: imputation and weighting adjustments (see, for example, Kalton, 1986). Brick (2013) discusses the advantages and disadvantages of each approach.

The imputation approach keeps the second wave nonrespondents in the analytic file, imputing their missing Wave 2 responses based on their baseline data. Performing these imputations in an effective way that does not distort relationships between items on a cross-sectional or longitudinal basis is the major challenge with this approach (see Brion, 2012, for a discussion of biases that can result from mass imputation). Until recently, this challenge has made the

weighting approach preferable to imputation. However, recent developments in imputation theory and software have made the imputation approach more attractive when the number of units to be imputed is relatively small, especially since, with it, the baseline weights of interview respondents are not altered for longitudinal analyses (Judkins et al., 2007).

The traditional approach to compensate for wave nonresponse has been by a weighting adjustment. Särndal and Swensson (1987) discuss approaching nonresponse adjustments as analogous to two-phase sampling, which would allow the baseline interview data from all adults to be used in constructing weights for adults who participate at Wave 2. Because there is so much information from the baseline about Wave 2 respondents and nonrespondents, the challenges with this approach are in the selection of auxiliary variables to be used for making the nonresponse adjustments and in determining the form of adjustment to use. A variety of methods, such as CHAID (Chi-squared Automatic Interaction Detector), logistic modeling of response propensity, and data mining, exists for determining the weighting classes. For example, Rizzo, Kalton, and Brick (1996) describe analyses they performed under a contract with the Census Bureau to examine these issues for handling panel attrition in the Survey of Income and Program Participation. Phipps and Toth (2012) contrast data mining approaches with logistic modeling for determining weighting adjustments.

A complication arises in later waves if a respondent misses one wave but returns to the cohort in the following wave. With the imputation approach, the imputed values for the missing wave should be made consistent with the responses for the adjacent waves. With the weighting approach, those missing a wave can be incorporated in cross-sectional estimates for the later wave, but they will not provide data for longitudinal analyses involving the missing wave. A possible compromise approach is to apply weighting adjustments to account for second-wave nonrespondents for analyses at that time, but then later to impute responses for those nonrespondents at the second wave who respond at the third wave, incorporating both baseline and Wave 3 responses in the imputation model. This approach then makes use of the high quality information available for the missing Wave 2 responses (Brick, 2013, p. 274).

Analyses will be conducted to assess the sensitivity of estimates to the variables used in the weighting or imputation procedure. Micklewright et al. (2012) describe methods that may be used to assess sensitivity to nonresponse adjustments in a survey in which large amounts of administrative data are available for the respondents and nonrespondents. These methods can be adapted for studying possible bias from attrition in the PATH Study, where interview data are available for the baseline respondents.

## **B.2f Expected Levels of Precision of the PATH Study**

The PATH Study is designed to produce reliable estimates of between-person differences and within-person changes in tobacco-related attitudes, behaviors, and health conditions among various population subgroups and over time. Many characteristics of interest are dichotomous, having “yes” or “no” outcomes. The percentage of “yes” responses is denoted by  $p$  and represents the prevalence estimate for a particular characteristic (e.g., cigarette smoking). Based on past research and cumulative professional expertise, the majority of characteristics measured in the PATH Study are expected to have magnitudes of prevalence exceeding 10 percent, while the expected magnitude of a few characteristics (such as initiation of tobacco use) will lie between 1 and 5 percent.

One measure of the precision associated with cross-sectional prevalence rates is relative standard error (RSE), defined as the standard error divided by the prevalence estimate and expressed as a percentage. More specifically,  $RSE(p) = 100 * \text{Standard Error}(p) / p$ , where the standard error is given by the square root of the variance of the estimate, taking into account the complex sample design of the PATH Study. A measure of power associated with longitudinal analyses of change in prevalence rates is the minimum detectable absolute difference (MDAD; see Lipsey, 1990). Herein, the MDADs represent the smallest change (up or down) from a given baseline prevalence rate that can be detected with 80 percent power using a two-sided test for equality of proportions at the 5 percent level of significance, taking into account the complex sample design of the study. The impact of the various complex features of the sample design on variances, and therefore on RSEs and MDADs, is reflected through inflation

factors called design effects (DEFFs). The extent to which these design effects exceed one indicates the extent to which the variance of an estimate based on the complex sample design is greater than the corresponding variance based on a simple random sample (SRS) design. Several key features of the PATH Study sampling design contribute to the overall design effect.

The first feature is the clustering at both the PSU and segment levels. In general, for a fixed sample size, the greater the number of units to be sampled per cluster, and the more homogeneous the sampling units are with respect to a characteristic of interest within clusters, the greater the DEFF and hence the inflation in the variance (resulting in decreased precision). The level of homogeneity within a cluster is reflected through two types of intraclass correlations:  $\rho_1$  for PSUs and  $\rho_2$  for segments. Note that  $\rho_1$  and  $\rho_2$  will vary in value for different characteristics of interest. The expected standard errors for prevalence estimates for the PATH Study have been calculated taking into account the contributions due to clustering at both the PSU and segment levels under the assumptions that the intraclass correlations ( $\rho_1, \rho_2$ ) are (.01, .05). These values were based on estimates taken from various sources in the survey research literature (see, for example, Guilliford et al. [1999] and Thompson et al. [2012]). The calculations reflect that “certainty PSUs” are in fact strata, not PSUs; therefore, there is no contribution to the variance from clustering at the PSU level for these PSUs. Thirty-five of the 156 PSUs selected are certainties, representing 24 percent of the U.S. population.

A second feature of the PATH Study design that contributes to the overall sampling variability is the baseline sampling of adults with different selection probabilities according to their age, race, and tobacco usage (the latter both as reported by the household screener respondent and as self-reported by the adult at the second phase of screening). The unequal weighting DEFFs due to this feature of the sample design are expected to range from 1.00 to 1.67, depending on the domain of interest. For analyses that combine all adult respondents, this component of the unequal weighting DEFF is expected to be approximately 1.81.

The third feature of the PATH Study design that contributes to the overall sampling variability is the restriction that no more than two adults be sampled

from a participating household. This requirement contributes to the variability of weights because adults in some multi-person households are sampled at lower rates than persons of the same age, race, and tobacco usage group in single- or two-person households. The unequal weighting DEFFs due to this feature of the sample design are expected to range from 1.00 to 1.02, depending on the domain of interest. For analyses that combine all adult respondents, this component of the unequal weighting DEFF is expected to be negligible (i.e., approximately equal to 1). Note that for analyses of subgroups of race, say by age or sex, these DEFFs will diminish, because generally fewer members of the subgroups will contribute to the clustering effect.

Estimates of precision and power for the PATH Study at Wave 2 were calculated taking into account the DEFFs resulting from the three previously-described sample design features. These estimates are shown in Tables B-5 and B-6, for adults and youth, respectively. The projected RSEs are for a generic statistic estimating a prevalence rate of 15 percent (such as the percentage of the adult population who are every day cigarette smokers). The MDADs are for a generic statistic estimating change from a baseline prevalence rate of 10 percent (such as any non-cigarette tobacco use). Both the RSEs and MDADs presented here are for illustrative purposes.

In Tables B-5 and B-6, the RSEs are for cross-sectional estimates at Wave 2 and the MDADs are for a change from baseline to Wave 2. The subgroups of interest are defined in terms of tobacco-related behaviors, which are subject to change over time. This presents a challenge when trying to estimate the subgroup sample sizes in future waves of the PATH Study, particularly given the recent expansion of tobacco products on the market. Over time, participants sampled as youth will become young adults and those sampled as young adults (18 to 24 years of age) will move into the older age group. As a result, variation in weights among members of most subgroups will increase, and it is necessary to inflate the assumed values of the DEFFs that are due to unequal weighting. It is not possible to predict the precise inflation factor for each subgroup given the complication of unknown, future rates of switching, substituting, or multiple tobacco product usage. For these reasons, one inflation factor of 1.03 was estimated for each follow-up wave and applied to all subgroups, and the estimates of cross-sectional precision and of detectable changes across waves are presented for a small

number of subgroups (i.e., those for which the estimates are expected to be fairly robust to the assumptions made). As a consequence, the estimates herein should be interpreted with caution.

Table B-5. Adult sample sizes, relative standard errors (RSEs), and minimum detectable absolute differences (MDADs) at Wave 2\*

Group	Wave 2 sample size	RSE on 15% item	MDAD on 10% item
<b>All adults</b>	<b>29,103</b>	<b>2.7</b>	<b>0.5</b>
Wide-net users	20,954	3.0	0.5
Current and experimental users	16,983	3.2	0.5
Current users	15,150	3.3	0.5
Menthol smokers	4,220	5.3	0.8
Dual (smokers and smokeless tobacco users)	1,112	9.8	1.4
Daily users	10,185	3.7	0.6
Less-than-daily users	4,966	5.0	0.7
Current non-users under wide-net definition	8,149	4.1	0.6
Urine sample providers	11,373	3.6	0.7
<b>Adults ages 18-24</b>	<b>8,180</b>	<b>4.1</b>	<b>0.9</b>

\*As indicated in the footnotes to Table B-3, 85 percent of 18 to 24 year-old baseline respondents are expected to complete the Wave 2 interview, while the corresponding estimate for older adults is 86 percent.

Table B-6. Youth sample sizes, relative standard errors (RSEs), and minimum detectable absolute differences (MDADs) at Wave 2 under assumption of 90% retention rate\*

Group	Wave 2 sample size	RSE on 15% item	MDAD on 10% item
<b>All youth</b>	<b>12,642</b>	<b>2.9</b>	<b>0.7</b>
Current users	1,149	7.3	1.8
Current smokers	626	9.7	2.4
Menthol smokers	373	12.4	3.0
Experimenters	1,528	6.4	1.6
Never smokers	10,631	3.0	0.8
Susceptible never smokers	2,241	5.4	1.3
Never users	7,083	3.5	0.9
<b>Youth ages 12 to 13</b>	<b>4,212</b>	<b>4.2</b>	<b>1.5</b>
Current users	70	28.5	10.4
Current smokers	28	44.9	16.3
Menthol smokers	19	54.9	20.0
Experimenters	205	16.7	6.1
Never smokers	3,947	4.3	1.6
Susceptible never smokers	549	10.3	3.8
Never users	2,666	5.0	1.8
<b>Youth ages 14 to 17</b>	<b>8,430</b>	<b>3.3</b>	<b>0.9</b>
Current users	1,079	7.5	2.1
Current smokers	599	9.9	2.8
Menthol smokers	354	12.7	3.6
Experimenters	1,323	6.8	1.9
Never smokers	6,685	3.5	1.0
Susceptible never smokers	1,692	6.1	1.7
Never users	4,417	4.1	1.2

\*As indicated in Table B-3, the 90 percent retention/response rate is the percentage of youth baseline participants completing the Wave 2 interview.

The large sample of adult tobacco users at Wave 2 will allow analyses for many user subgroups as well as for persons who are considered at risk for becoming tobacco users. Table B-5 highlights subgroups of potential analytic interest by breaking out sample sizes and measures of precision and power for tobacco users, menthol cigarette smokers, users of both smoked and smokeless tobacco, daily/non-daily tobacco users, and young adults (18 to 24 years of age). In addition, the RSE and MDAD are shown for the sample of adults expected to provide urine specimens at Wave 2. The subgroup sample sizes for the different categories of tobacco users were estimated using data from replicate 1 (see

footnote 1 for the definition and discussion of this term) of the PATH Study. The “current user” definition from Section B.1b was applied in estimating sample sizes for menthol smokers, dual and daily user, and less-than-daily user groups; the “wide net” definition was used to estimate the sample sizes for nonuser groups. These are the definitions that give the smallest sample size, and hence the largest RSEs and MDADs, for each of these groups. The estimated RSEs and MDADs for another definition of tobacco user will be smaller than those displayed in the tables. For both RSEs and the MDADs, smaller is better. The RSEs for a 15 percent prevalence rate are at or below 5 percent for most subgroups shown. The MDADs for a 10 percent baseline prevalence rate are mostly below 1 percentage point indicating that a one-year change of 1 percentage point or less can be reliably detected for the subgroups shown. With the exception of the group who provide urine specimens, the adult sample sizes considered in this section are based on estimates for completed Wave 2 interviews; therefore, the estimates of precision and power apply to projected estimates of tobacco and health outcomes collected with the Wave 2 instruments. As is the case for all the estimates presented in this section, it is expected that precision and power will be reduced for finer divisions of the subgroups (e.g., by gender).

The initial sample of 14,050 youth at baseline was intended both to replenish the adult cohort in future waves of the PATH Study and to provide sufficient power for analyses of youth subgroups. Table B-6 shows Wave 2 sample sizes and measures of precision and power for the youth sample overall and by subgroups of possible interest: tobacco users, cigarette smokers, menthol cigarette smokers, “experimenters,” never smokers, susceptible never smokers, and never users of tobacco; the same statistics are shown for each of these subgroups among 12-to-13 year-olds and among 14- to-17 year-olds. Subgroup sample sizes were estimated using data from replicate 1 of the PATH Study. For youth, current smokers were defined as youth who have smoked a cigarette within the last 30 days, and current users were youth who have used any tobacco product within the last 30 days. Experimenters were defined as youth who have ever smoked any cigarette, even one or two puffs, but fewer than 100 cigarettes. Susceptibility to initiate cigarette smoking among never smokers was defined as providing any response other than “definitely not” to at least one of the questions: “Do you think that you will try a cigarette soon?”, “Do you think you will smoke a cigarette

anytime during the next year?" or "If one of your best friends offered you a cigarette, would you smoke it?"

Overall, there are large samples in many of the subgroups of interest. For example, there are approximately 7,080 never users for whom tobacco use initiation rates will be tracked. Tobacco cessation is more of an issue in the older adolescent group (ages 14 to 17) because there are more tobacco users in that age group than among youth ages 12 to 13, and there are about 1,080 tobacco users and 600 cigarette smokers whose quitting behavior over time will be monitored. The smallest subgroup presented in Table B-6 that may be of interest is menthol smokers; for example, if regulatory action related to menthol cigarettes were undertaken, these youth might respond by quitting, switching brands, or switching to other forms of tobacco use. For some of the subgroups with very small sample sizes, such as menthol smokers 12 to 13 years of age, precision is low. The PATH Study is expected to provide sufficient precision for studying subgroups of tobacco users/experimenters and nonusers among the 12-to-13 year olds; however, for subgroups with very small sample sizes, the age groups will be combined (e.g., 12- to 17-year olds) to produce estimates with higher precision by type of tobacco use.

The RSEs for a 15 percent prevalence rate among youth 12-to-17 years of age are below 13 percent for most subgroups, and at or below 7 percent for more than half of them. Among all youth 12-to-17 years of age, the sample size overall and in each of the subgroups except menthol smokers is sufficient to detect a one-year change of 2.5 percentage points in a 10 percent baseline behavior overall. This is an important threshold because measures of quitting, initiation, and non-cigarette tobacco use tend to be in this 10 percent range (depending on the definitions used); a statistically significant increase (or decrease) of two and a half percent would be pertinent information for FDA in meeting its regulatory authorities under the TCA.

## **B.3 Methods to Maximize Response Rates and Deal with Nonresponse**

For Wave 2 and subsequent annual follow-up waves, the PATH Study will focus on maintaining contact with respondents and maximizing their retention in the study. The methods used by the study to meet these objectives include: (1) tracking participants and tracing them, as needed; (2) maintaining a sufficiently large field interviewer workforce located near the selected PSUs; (3) implementing robust interviewer training and quality control procedures; (4) interviewing in Spanish as well as in English; (5) communicating with participants by mail and telephone in advance of in-home data and biospecimen collection visits; (6) continuing to emphasize the importance of biospecimen collections to field interviewers and respondents; and (7) attempting to convert refusals from adults.

Specific to the PATH Study, OMB's terms of clearance in approving the baseline require NIDA and FDA to report to OMB the predicted response rates associated with the baseline (for screening, interview completion, and biospecimen collection), the results of nonresponse analysis and the plan for future statistical analyses, and the implications of the response rates and nonresponse bias for the types of conclusions that can be drawn from this study. Section B.4 summarizes the results of an interim report to OMB on these topics; the interim report is provided in Attachment 21.

### **B.3a Maintaining Participant Engagement and Tracking**

The PATH Study seeks to maintain respondent engagement as well as track respondents so they can be contacted for follow-up data and biospecimen collection. The following activities are planned for Wave 2 to maintain respondent engagement:

- Mail a thank you card after the baseline interview and a birthday card to adult SPs and the parents of youth SPs;

- Mail a follow-up letter to the parents of shadow youth at approximately 6 months after the baseline contact and telephone the parents at approximately 1 year after the contact;
- Visit respondents who have moved up to 100 miles from a study PSU,<sup>14</sup> and
- When feasible, attempt to visit respondents who have moved more than 100 miles from a study PSU.

Ongoing tracking of study respondents is essential to longitudinal cohort studies for purposes of cohort retention and follow up. Management of participant tracking and tracing activities by the PATH Study is through a centralized Home Management System (HMS). This component of the study management system houses the database of contact information, and it provides for real-time access in the field and at the home office to the most current information available. PATH Study staff involved in tracking and tracing activities provide updates to the HMS, and supervisors generate reports from the HMS to monitor progress in the field and identify the need for potential corrective actions.

The centralized HMS tool facilitates routine tracking steps that help to minimize the number of cases requiring intensive tracing. These steps include:

- **Collect contact information at baseline for tracing references.** At baseline, respondents are asked for the names, addresses, and telephone numbers of two people who do not live in the same household and can serve as tracing references for how to reach the respondent. Given that a sizeable percentage of respondents are young adults, respondents are asked for additional information that may help to locate them (e.g., recent college attended).
- **Use interim contacts to determine if contact information has changed or if tracing is needed.** Contacts by mail ask respondents to report any address changes, and the study provides a number of easy ways this can be done, including visiting the study website, calling a toll-free number, or sending updated information via mail. The PATH

---

<sup>14</sup> The proportion of SPs who move annually beyond 100 miles from any of the PSUs is estimated to be less than 1 percent.

Study also mails materials to respondents stamped “return service requested,” requesting new address information for people who may have moved. In addition to supporting tracing, these interim contacts help to maintain respondent motivation to cooperate and continue engagement with the study. PATH Study respondents are offered an incentive (\$5) as a thank you for updating their contact information on up to two occasions during the year, for a total of \$10.

- **At each in-person visit, update contact information.** During household visits for each follow-up wave, the field interviewers update contact information on the respondent as well as on relatives or persons not living in the household who can serve as references on where to locate the respondent.

The PATH Study has a robust, systematic approach for tracing and locating respondents who may be lost at the Wave 2 follow-up. If current occupants of the last known address are unable to guide the field interviewer to a respondent’s whereabouts, the field interviewer implements the first line of tracing using readily available information, including the respondent’s last known telephone number(s), tracing references, directory assistance, and neighbors to try to locate the respondent. If unsuccessful, the case is sent to the PATH Study home office for the second line of tracing, which is more intensive. An expert team of tracers at the home office follows established protocols to trace and locate PATH Study respondents. These protocols include the following.

- **Lexis Nexis.** This database, compiled from public records, can return respondent address histories and telephone numbers. Submissions are made at least quarterly, and the tracers review and follow up on the results.
- **Internet searches.** These searches include free and paid services. Examples of the services include online telephone directories and limited public information records.
- **In-person tracing.** As the need arises and as resources permit, in-person tracing (i.e., “skip tracing”) may be used. This approach involves intensive in-person tracing at the respondent’s last known addresses and in his/her old neighborhoods to identify contact information or current location; in-person tracing differs from the first line of tracing

by using specialists who develop leads that extend beyond those based on readily available information. Given its expense on a per case basis, in-person tracing is used rarely, after exhausting other approaches.

### **B.3b Wave 2 Data and Biospecimen Collection**

In an effort to minimize attrition and maximize response rates in advance of Wave 2 data and biospecimen collection (as well as throughout each follow-up wave), the PATH Study has a team of highly experienced field interviewers and field supervisors ready to work all cases thoroughly. These field interviewers are strategically located within or in close proximity to PSUs to help expedite visits to sampled persons' homes, ensuring that they and other field staff are familiar with the communities within which their assigned cases are located. Field interviewers are trained in effective techniques to gain respondent cooperation through refusal aversion and conversion.

The PATH Study uses several tools and approaches to address nonresponse and maximize response rates in addition to the respondent incentives described in Section A.9. These tools and approaches include the following.

- The interviews are conducted in English and Spanish; all of the instruments are translated into Spanish, and bilingual field staff administer them.
- Materials for study respondents are designed to be informative and to encourage participation; all of the materials are translated into Spanish. These include follow-up or reminder letters that are sent 3 months prior to the planned interview date to inform adult SPs and the parents of youth SPs about the planned Wave 2 (Attachment 9). The letters remind recipients about the PATH Study's objectives, how its data will be used, why the study is important, and why the study is interested in including tobacco users and non-users.
- Respondents can easily access information about the PATH Study through the PATH Study website and a toll-free respondent telephone

call line dedicated to answering respondents' questions and verifying the credibility of the study.

Approximately 1 month prior to the planned interview date, field interviewers telephone the adults SPs and parents of youth SPs. This call is intended to reestablish direct contact, answer questions about the study, and make an appointment for the in-person visit. As needed (e.g., adult SP does not have a telephone), the field interviewers make the first direct contact in-person.

Tailored letters are used with reluctant respondents/sample persons and with selected units located in limited-access situations (doorperson buildings, gated communities, etc.), which may be sent via FedEx or priority mail to reinforce the perceived importance of participation. (See Attachment 19 for an example of a refusal letter.)

Additional tools and approaches will be used by the PATH Study to help maximize the biospecimen response rates for Wave 2, including the following.

- **Ensure that Interviewers are “On Board.”** The PATH Study continues to hire and train interviewers who understand the importance of collecting biospecimens as part of this research effort. Early in the selection process, candidates are required to view a short video that highlights this requirement and the importance of being comfortable with carrying it out.
- **Phase the Consent for Biospecimens.** For youth SPs who age into the adult cohort, the PATH Study presents information to respondents in phases to help minimize the amount of information to be simultaneously considered before consenting. This approach includes providing information about the interview immediately prior to obtaining consent for the interview; providing information on biospecimen collection shortly before obtaining consent for biospecimen collection, etc. Moreover, because biospecimen collection follows completion of the interview, this approach also allows additional time for the development of rapport, trust, and comfort between the interviewer and the respondent, which positively influence consent to provide the biospecimens.

- **Present the Biospecimens in a Positive Light.** Based on an effective approach used by the National Health and Nutrition Examination Surveys (NHANES), the PATH Study uses nicely-formatted consent pamphlets with messages that emphasize the importance of the respondent's contributions of biospecimens to the study's scientific success.
- **Enhance Training of Interviewers.** The PATH Study continues to provide extensive interviewer training on collecting biospecimens, including home-study training and practice in requesting consent and averting refusals. With classroom and home-study training and additional practice sessions, interviewers are able to gain proficiency and comfort with the study protocol, including obtaining consent, averting refusals, and collecting biospecimens.
- **Equip Interviewers with Refusal Conversion Tool.** The PATH Study continues to use computer-assisted personal interviewing (CAPI) screens that, in real time, point interviewers to tailored responses to types of reasons respondents give for biospecimen refusals. Having these available at the moment they are needed can improve the interviewer's ability to quickly allay respondent concerns about providing biospecimens.
- **Streamline Biospecimen-Collection Procedures.** The PATH Study will continue its procedure of having the field interviewer collect a urine sample at the time of the interview for participants providing urine at Wave 2. Rapport that develops between the interviewer and respondent may also have a positive influence on the respondent's willingness to provide the biospecimens. As noted, these participants will be the youth who age into the adult cohort at Wave 2 and a subsample of adults who provided a urine sample at baseline.
- **Enhance Quality Control.** The PATH Study's data collection quality control procedures include closely monitoring interviewer-by-interviewer consent and collection rates for biospecimens, using computer-assisted recorded interviewing (CARI) to monitor interviewer performance on the consent and collection tasks, and providing rapid feedback to interviewers and refresher training to maximize performance.

A web-based Supervisor Management System (SMS) allows field supervisors to monitor each field interviewer's work and help in the development of strategies to address nonresponse. These strategies may include reassigning difficult or reluctant cases among local field interviewers; and using specially-trained, traveling field interviewers with experience in refusal conversion.

Data collection efforts also follow a phased approach that anticipates refusal conversion efforts. In this approach, samples of SPs are released to field interviewers in sets every few months; the timing of these releases approximates the anniversary of when the baseline interviews were completed. Hence, closing out cases from an earlier set is not necessary before releasing cases in a new set, thus allowing additional time to complete challenging cases. Further, the number of cases assigned to interviewers is expected to be lowest during later periods in the data wave, thereby ensuring interviewers have additional time in those periods to complete open cases remaining from an earlier period. Front-loading the sample release in this manner allows field interviewers the opportunity to implement the full contact strategy, including nonresponse conversion as needed.

Adjustments will be performed as necessary for non-interviews that cannot be converted using the procedures described in Section B.2. The specific procedure selected ensures the accuracy of resulting estimators and the suitability of the compensated data set for addressing the major objectives of the study.

The Wave 2 response rate is estimated to be 85 percent for young adults (ages 18 to 24 at baseline), 86 percent for older adults, and 90 percent for youth. (See Section B.1 for a discussion of these and other estimated response rates.) These rates are calculated as the number of respondents divided by the number of eligible sample persons. For Wave 2, ineligible persons include persons under the age of 12 years and persons responding at baseline who die before the Wave 2 data and biospecimen collection begins.

## **B.4 Test of Procedures or Methods to Be Undertaken**

The PATH Study baseline data and biospecimen collection, which is currently underway, serves as an informal test of many of the methods and materials planned for Wave 2. This is reasonable, because many of the baseline and Wave 2 methods and materials are the same. For example, for the cohort movers (youth SPs who age into the adult cohort and shadow youth who age into the youth cohort), the Wave 2 consent, parental permission, and youth assent procedures and methods closely resemble those used in baseline. Also, the Wave 2 methods for administering the ACASI and CAPI instruments, collecting biospecimens, and paying incentives for all types of participants are similar to the methods used for baseline.

In addition to this informal test, the PATH Study developed the afore-mentioned interim report based on the first 5 months of baseline data and biospecimen collection. Findings in the interim 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. Response rates based on 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 the baseline wave from September 12, 2013 to February 26, 2014.

The interim report is provided in Attachment 21. Some of its findings are highlighted in the remainder of this section.

### **B.4a Predicted Response Rates**

Predicted response rates, including response rates weighted with inverse probabilities of selection (IPS), were computed for the three data collections and

three biospecimen collections.<sup>15</sup> The predicted response rates for the collections vary on how they compare to the “best-case” and “worst-case” scenarios. (Please see the introduction and reference to Attachment 22 in the interim report for more information on these scenarios.)

- The weighted predicted response rates for the Household Screener (57.1%) and Adult Extended Interview (75.7%) are lower than the best-case scenarios for the full sample, but they exceed the worst-case scenarios. The weighted predicted response rate for the Youth Interview (81.2%) exceeds the best-case scenario.
- The weighted predicted response rates for the biospecimen collections—buccal cells (69.0%), urine (61.8%), and blood (36.9%)—are lower than the best-case scenarios. The weighted predicted response rates for the buccal cell and blood collections are slightly lower than the worst-case scenarios.
- As discussed further in the report, differences among tobacco use status and demographic subgroups on predicted response rates for the collections are generally modest.

## **B.4b Nonresponse Bias Analysis**

PATH Study IPS-weighted estimates were compared with estimates from other studies. Most of the PATH Study estimates are consistent with those from other studies.

- The PATH Study estimates of percentage of single-person households are lower than those in the American Community Survey (ACS).
- Hispanics, adults 25 to 44 years old, and adults with higher education levels are somewhat over-represented in the PATH Study estimates for some specific data or biospecimen collections.

---

<sup>15</sup> During the baseline wave, in addition to urine and blood samples, buccal cells were collected from adult participants. Due to budgetary constraints, it was necessary to discontinue buccal cell collection before the end of the wave.

- The PATH Study’s estimates of adult cigarette smoking are in line with those from other studies. Its estimates of youth cigarette smoking are toward the low end of the range of estimates found for other studies.<sup>16</sup>

#### **B.4c Approach to Address Nonresponse**

The statistical approach to address nonresponse is to adjust the IPS weights to account for nonrespondents. This approach was successful in correcting for nonresponse bias on characteristics measured in the ACS.

- Applying the adjusted IPS weights to the predictor sample reduces the discrepancy between the PATH Study estimates and ACS estimates on demographic characteristics.
- Estimates of adult cigarette smoking using the adjusted weights remain in line with those from other surveys.
- Estimates of youth smoking using the adjusted weights are close to the estimates using the IPS weights and are still at the low end compared with other surveys.

### **B.5 Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data**

A list of individuals who consulted on statistical aspects of the PATH Study design and will collect and/or analyze the PATH Study data is included in Attachment 22.

---

<sup>16</sup> Estimates of adult cigarette smoking from the PATH Study were compared with estimates from 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); and the National Survey on Drug Use and Health, 2012 (NSDUH). Estimates of youth cigarette smoking from the PATH Study were compared with estimates from NHANES, NSDUH, and the National Youth Tobacco Survey, 2012 (NYTS). Results from the 2013 Youth Risk Behavior Surveillance Study were released on June 13, 2014 (Kann et al., 2014), after the Interim Report was prepared; these indicate that youth cigarette smoking dropped from 7 to 13 percent for various subgroups between 2011 and 2013.

## References

- Brick, J.M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*, 29, 329–374 (with discussion).
- Brion, P. (2012). Discussion of “Calibrated Bayes, an alternative inferential paradigm for official statistics.” *Journal of Official Statistics*, 28, 341–347.
- Contoyannis, P., Jones, A. M., and Rice, N. (2004). The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics*, 19, 473–503.
- De Waal, T., Pannekoek, J., and Scholtus, S. (2013). *Handbook of Statistical Data Editing and Imputation*. Hoboken, NJ: Wiley.
- Guilliford, M.C., Ukoumunne, O.C., and Chinn, S. (1999). Components of variance and intraclass correlations for the design of community-based surveys and intervention studies. *American Journal of Epidemiology*, 149(9), 876–883.
- Judkins, D., Krenzke, T., Piesse, A., Fan, Z., and Haung, W.C. (2007). Preservation of skip patterns and covariate structure through semi-parametric whole questionnaire imputation. *American Statistical Association Proceedings of the Survey Research Methods Section*: 3211–3218.
- Kalton, G. (1986). Handling wave nonresponse in panel surveys. *Journal of Official Statistics*, 2, 303–314.
- Kann, L., Kinchen, S., Shanklin, S.L., et al. (2014). Youth Risk Behavior Surveillance – United States, 2013. *MMWR* 63(4), 1–168.
- Kashihara, D. and Ezzati-Rice, T. (2004). Characteristics of survey attrition in the household component of the Medical Expenditure Panel Survey. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 3758–3765.
- Lipsey, M. (1990). *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: Sage.
- Lynn, P. (2006). Quality Profile: British Household Panel Study. Essex, UK: Institute for Social and Economic Research. Accessed 05/08/2014 from

<https://www.iser.essex.ac.uk/files/bhps/quality-profiles/BHPS-QP-01-03-06-v2.pdf>.

- Micklewright, J., Schnepf, S., and Skinner, C. (2012). Non-response biases in surveys of schoolchildren: The case of the English Programme for International Student Assessment (PISA) samples. *Journal of the Royal Statistical Society, Series A*, 175, 915–938.
- National Institutes of Health (2010). Alcohol use and alcohol use disorders in the United States, a 3-year follow-up. U.S. Alcohol Epidemiologic Data Reference Manual, Volume 8, Number 2. Accessed 05/27/2014 from [pubs.niaaa.nih.gov/publications/NESARC\\_DRM2/NESARC2DRM.pdf](https://pubs.niaaa.nih.gov/publications/NESARC_DRM2/NESARC2DRM.pdf).
- National Longitudinal Survey of Youth, 1997 (2014). Retention & Reasons for Non-Interview. Accessed 05/22/2014 from <https://www.nlsinfo.org/content/cohorts/nlsy97/intro-to-the-sample/retention-reasons-non-interview/page/0/0/#retention>.
- National Research Council (2014). *Nonresponse in Social Science Surveys: A Research Agenda*. Washington, D.C.: National Academies Press.
- Phipps, P. and Toth, D. (2012). Analyzing establishment nonresponse using an interpretable regression tree model with linked administrative data. *The Annals of Applied Statistics*, 6, 772–794.
- Rizzo, L., Kalton, G., and Brick, J.M. (1996). A comparison of some weighting adjustment methods for panel nonresponse. *Survey Methodology*, 22(1), 43–53.
- Thompson, D.M., Fernald, D.H., and Mold, J.W. (2012). Intraclass correlation coefficients typical of cluster-randomized studies. *Annals of Family Medicine*, 10, 235–240.
- Van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Boca Raton, FL: CRC Press.