

**Supporting Statement B for:**

**The Population Assessment of  
Tobacco and Health (PATH) Study (NIDA)**

**—**

**Fourth Wave of Data Collection**

**OMB No. 0925-0664, Expiration Date:  
8/31/2018**

**June 27, 2016**

**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

<b>B.</b>	<b>Collections of Information Employing Statistical Methods</b>	
B.1	Respondent Universe and Sampling Methods.....	1
B.2	Procedures for the Collection of Information.....	12
B.3	Methods to Maximize Response Rates and Deal with Nonresponse.....	32
B.4	Test of Procedures or Methods to be Undertaken.....	37
B.5	Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data.....	41
	References .....	42

## LIST OF ATTACHMENTS

- Attachment 1. NIDA's Data Collection Authority
- Attachment 2. PATH Study Data Collection Instruments
- Attachment 3. Crosswalk of PATH Study Objectives, Data Sources, Domains, and Analysis
- Attachment 4. Summary of Changes to Instruments for Wave 4
- Attachment 5. Sample Conceptual Models
- Attachment 6. Additional Information on Biospecimens
- Attachment 7. PATH Study Privacy Impact Assessments
- Attachment 8. Key Design Features of National Tobacco Surveys
- Attachment 9. Follow-up, Retention, and Tracking Materials



## **LIST OF ATTACHMENTS (continued)**

Attachment 10. List of Consultants

Attachment 11. Consent Materials

Attachment 12. Certificate of Confidentiality

Attachment 13. Assurance of Confidentiality

Attachment 14. IRB Approval Memo

Attachment 15. NIH Privacy Act Systems of Record Notice

Attachment 16. Procedures for Keeping Data Confidential

Attachment 17. ICPSR Data Use Agreement

Attachment 18. Field Data Collection Materials

Attachment 19. Sample Analysis Plans

Attachment 20. PATH Study Interim Report

Attachment 21. List of Statistical Consultants

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

This document describes the statistical methods planned for Wave 4 of the PATH Study. Section B.1 describes the target populations of the PATH Study, the respondent universe, and the actual or expected sample distributions across the PATH Study waves based on such characteristics as age, tobacco-use, and race-ethnicity. It also discusses the PATH Study sample design and the expected response rates for Wave 4. Section B.2 describes data collection, weighting, and estimation procedures and expected precision estimates for the analyses of various domains of interest. Section B.3 describes procedures to maximize participation and retention of the PATH Study respondents. Section B.4 discusses plans for evaluating data collection procedures, including a discussion of nonresponse bias. The final section, Section B.5, presents a list of the statistical consultants contributing to the PATH Study.

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

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

The PATH Study is an ongoing longitudinal study in which all participants are followed as long as they are living in the U.S. with the exception of those who are incarcerated. The original sample included a “shadow sample” of youth ages 9 to 11 years old to support the addition of a sample of youth who age into the youth cohort at the subsequent three annual waves of data collection. At each wave of the study, shadow youth who have aged into the youth age range are added to the target population from that wave on. Also, youth who reach the age of 18 at a given wave are treated as adults from that wave on. The PATH Study design is constructed to serve the study’s important research (analytic) interests, including both longitudinal and cross-sectional analyses.

The original cohort (referred to here as the “Wave 1” cohort) of the PATH Study will continue to be interviewed at Wave 4. Moreover, additional sample will be added to the study at Wave 4 to provide a shadow sample for Wave 5 and to compensate for panel attrition by increasing the sample size to at least the Wave 1 level. The combination of the continuing initial cohort in the civilian, household population at Wave 4 with the replenishment sample will comprise a new cohort (referred to here as the “Wave 4” cohort) that will similarly be followed in future waves. The target population for this new cohort consists of adults and youth who are members of the U.S. resident civilian, household (noninstitutionalized) population at the time of Wave 4.

The Wave 4 respondents from the sample selected at Wave 1 will allow the PATH Study to continue estimating longitudinal change within persons over time. A separate set of weights will be developed to permit cross-sectional estimation using Wave 4 participants from both the Wave 1 sample and the replenishment sample for Wave 4, that is, including the continuing, aged-up, and newly sampled members of the Wave 4 sample.<sup>1</sup>

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

Estimates of the population numbers in the youth respondent universe at Wave 1 as well as youth respondent sample sizes for Waves 1 to 4 of the PATH Study are shown in the second row of Table B-1. The youth respondent universe is based on estimated population counts of persons eligible for the PATH Study from the 2013 American Community Survey (ACS). As shown in Table B-1, the number of completed interviews with youth 12 to 17 years old at Wave 1 is 13,651. After accounting for Wave 1 shadow sample<sup>2</sup> members

---

<sup>1</sup> Continuing participants are persons who participated in a previous wave in the same age cohort. Aged-up adults are persons who were previously interviewed as youth and are newly eligible for the adult cohort. Aged-up youths are persons who were previously tracked as “shadow youth” and are newly eligible for the youth cohort. New participants are persons who are selected from the Wave 4 replenishment sample.

<sup>2</sup> The “shadow sample” consists of children who were between the ages of 9 and 11 at the household’s Wave 1 interview. These children were not

who have turned 12, youth interviewed at Wave 1 who have become adults, and expected attrition among the remainder of the Wave 1 youth cohort, the expected number of Wave 1 cohort youth completing the Wave 4 youth survey is 10,365. With 4,693 youth respondents expected from the replenishment sample selected at Wave 4, the total estimated number of completed interviews with youth 12 to 17 years old at Wave 4 is 15,058. Estimated numbers for Waves 2 and 3 are also provided in Table B-1.

Population estimates of the adult respondent universe are shown in Table B-2, which presents estimated numbers by domain representing cross-classifications of age, race, and tobacco use as they were used for sampling purposes. There are many possible definitions of “tobacco user.” The one used here is known as the “wide net” definition of tobacco use, which classifies a person as a tobacco user if he/she satisfies at least one of the two following conditions: having smoked a cigarette, cigar, or pipe, or used smokeless tobacco in the past 30 days; and/or having 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.

In addition to Wave 1 population estimates, Table B-2 provides actual or estimated sample sizes for each of the first four waves of the PATH Study. For Wave 4, the expected numbers of respondents are provided for the continuing and replenishment samples separately as well as combined.

The respondent universe counts in the second column of Table B-2 were computed by estimating the total number of wide-net tobacco users and nonusers in the adult civilian population for each age/race domain from Wave 1 of the PATH Study.<sup>3</sup> The number of completed adult interviews at Wave 1 was 32,320, including 9,112 young adults (18 to 24 year olds) and

---

interviewed at Wave 1, but they will be enrolled in the youth cohort in a subsequent wave when they turn age 12.

<sup>3</sup> These estimates were calculated using the final adult weights, which were calibrated to population estimates from the 2013 ACS.

5,580 Blacks or African Americans (Black/AA).<sup>4</sup> At Wave 4, the corresponding estimates are 34,151 completed adult interviews, with 10,943 young adults and 5,941 Blacks or African Americans. These numbers account for both aging of the Wave 1 sample participants and expected attrition. The PATH Study will generate longitudinal data on a range of tobacco use behaviors within the cohort. Pending the availability of these data, the sample sizes presented in Table B-2 for Waves 3 and 4 are estimated using the wide-net tobacco use rates calculated from the Wave 1 sample.

Except for the number of youth in the shadow sample (i.e., 9 to 11 year olds selected at Wave 1 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 Wave 1, Wave 2, Wave 3, and Wave 4 *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 Wave 1. Power projections are provided later in Supporting Statement B for subgroups of potential analytic interest.

Table B-1. Numbers in the PATH Study youth and shadow youth respondent universe for Wave 1, as well as sample sizes for Waves 1 and 2, and estimated sample sizes for Waves 3 and 4

Group	Respondent universe	Wave 1 sample size	Wave 2 sample size	Estimated Wave 3 sample size	Estimated Wave 4 sample size from continuing sample (no shadow)	Estimated Wave 4 sample size from replenishment sample (only ages 10-11 for shadow)	Estimated Wave 4 sample size
Children 9-11	12,273,57	7,207	4,207	1,767	---	4,684	---

<sup>4</sup> Questions in the PATH Study instruments that collect data on race or ethnicity are 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).

(shadow sample)	5							
Youth 12-17	24,852,363	13,651	12,172	11,258	10,365	4,693	15,058	

Table B-2. Final domain totals within the PATH Study adult respondent universe, sample sizes for Waves 1 and 2, and estimated sample sizes for Waves 3 and 4 using the “wide net” definition of tobacco use (departures may arise in summary totals due to rounding)\*

Group	Estimated Wave 1 population totals by adult domain within the respondent universe	Wave 1 sample size	Wave 2 sample size	Estimated Wave 3 sample size	Estimated Wave 4 sample size from continuing sample	Estimated Wave 4 sample size from replenishment sample	Estimated Wave 4 sample size
18-24 Black/AA user	2,746,432	1,306	1,159	901	808	538	1,346
18-24 Black/AA non-user	2,263,913	510	491	648	721	110	831
18-24 non-Black/AA user	15,116,956	5,546	4,668	4,000	3,668	2,028	5,696
18-24 non-Black/AA non-user	10,601,252	1,750	1,858	2,383	2,692	378	3,070
25+ Black/AA user	10,583,809	2,559	2,265	2,058	1,968	591	2,559
25+ Black/AA non-user	14,970,662	1,205	1,114	948	898	307	1,205
25+ non-Black/AA user	60,671,050	13,675	11,482	10,711	10,129	3,546	13,675
25+ non-Black/AA non-user	119,737,510	5,769	5,338	4,400	4,114	1,655	5,769
All adults	236,693,582	32,750	28,375	26,048	24,998	9,152	34,151

\*Note that the Wave 2 “wide net” definition was slightly different from that for Wave 1, replacing respondent status for “has ever used an e-cigarette” with that for “has ever used an e-product.”

### B.1c Sample Design

The sample design for Wave 4 focuses on the blending of the replenishment sample selected at Wave 4 with the continuing sample begun at Wave 1. Wave 3 respondents will continue to be followed in Wave 4 as will most Wave 2 respondents who were nonrespondents to Wave 3. Some types of

earlier wave nonrespondents (e.g., hostile refusers<sup>5</sup>) will not be fielded in Wave 4 (see Section B.2a). These previous PATH Study participants fielded in Wave 4 form what is referred to as the Wave 1 cohort and represent the longitudinal population described in section B.1a on “Target Populations.”

In addition, a “replenishment” sample will be selected for fielding in Wave 4. The replenishment sample will be allocated across the same sample strata employed for Wave 1 so that, when the respondents from the replenishment sample are pooled with the Wave 4 respondents from the Wave 1 cohort, the corresponding “final” domains (representing the same subgroups as the sample strata but for respondents, and thus accounting for misclassification resulting from household screener assignments, mainly associated with tobacco use) will be increased to at least their Wave 1 levels.

The replenishment sample will be selected from members of the civilian, household population, while the Wave 1 cohort will include some Wave 4 respondents who are members of the military or the institutionalized population (e.g., nursing homes), as PATH Study respondents are followed over time as long as they are members of the U.S. resident population (i.e., those residing in the U.S., regardless of military or institutional status). Moreover, the replenishment sample will have been selected from a population that covers people who were not members of the civilian, household population at the time of Wave 1 (e.g., immigrants, those returning to civilian status from the military, and those returning to the noninstitutionalized population from jail). As a result, a new cohort, referred to here as the Wave 4 cohort, will be formed for analytic purposes through pooling the replenishment sample with the members of the Wave 1 cohort who are in the civilian, household population in Wave 4. The replenishment samples for the various sampling strata will be selected by probability sampling from a screening sample selected within the same PSUs and segments as the Wave 1 sample.

---

<sup>5</sup> Hostile refusers are persons who refuse requests for study participation in an angry or threatening manner.

The shadow sample allows the PATH Study to recruit youth under age 12 for future waves once they turn age 12. Thus, during Wave 1 sampling, 9 to 11 year olds were selected for participation in subsequent waves (Waves 4, 3, and 2, respectively). The PATH Study collects no interview data from the shadow sample; they are followed over the waves and invited to join the youth cohort when they become eligible. Waves 1 to 4 will have been carried out in four consecutive years. Wave 5 of the PATH Study will be a biannual, rather than annual wave, conducted two years after Wave 4. Thus, as part of the Wave 4 sampling, a sample of 10 and 11 year olds will be identified for participation in the youth interview for Wave 5 when they will be 12 and 13 year olds.

Ten and 11 year olds represent a small percentage of the U.S. population, but they will contribute a sizeable proportion (roughly one-third) of youth respondents for the next wave of the study.

The overall goals for the full sample to be interviewed at Wave 4 are to:

- Achieve a total Wave 4 sample size of at least 45,971 respondents with at least 32,320 adults and at least 13,651 youth;
- Provide the capability for longitudinal analyses, including estimates of behavioral and health measures for subgroup comparisons by tobacco user status, type of tobacco product, race/ethnicity, age, sex, and other characteristics using data associated with the Wave 1 cohort; and
- Provide the capability for cross-sectional analyses, including estimates of behavioral and health measures for subgroup comparisons by tobacco user status, type of tobacco product, race/ethnicity, age, sex, and other characteristics using data associated with the Wave 4 cohort.

The following sections describe the sample design. As background, an overview is provided of the sample design for Waves 1 to 3 that will have been implemented prior to fielding Wave 4. Then, a summary is presented of the key features of the sample design for the replenishment sample selected at Wave 4.

## ***Background: Overview of the Sample Design for Waves 1 to 3 and the Longitudinal Component of Wave 4***

The Wave 1 sample design served as the framework for fielding Waves 1 to 3 of the PATH Study, and will continue to serve as the basis for the longitudinal component of Wave 4. Briefly, the Wave 1 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) 166,088 addresses; and (4) 76,526 eligible sampled persons (SPs) within responding households occupying dwelling units (DUs) at sampled addresses. 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 Wave 1. 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 Wave 1 sample are described in Sections B.1c and B.1d of Supporting Statement B for Wave 1.

Primary Sampling Units (PSUs) consisting of one or more counties were formed covering all counties in the U.S. These PSUs were assigned to strata and 156 were randomly selected with probability proportionate to size (pps). Within sampled PSUs, segments, consisting of one or more Census blocks, were also randomly selected with probability proportional to size (pps). Address Based Sampling (ABS) was then used to select the vast majority of addresses to be fielded (94.5%) from the U.S. Postal Service's Computerized Delivery Sequence File (CDSF).

The CDSF provides high coverage of residential addresses in the U.S., except for some rural areas, which required traditional listing methods for sampling addresses in some segments. In addition, Address Coverage Enhancement (ACE) methods were used to include addresses that do not appear on the

CDSF (accounting for households where, for instance, the only mail delivery is through a P.O. Box). A little over 158,000 addresses were selected from the CDSF. Close to 1,800 sampled addresses (about one percent of the sample) were selected from traditionally listed addresses, and roughly 7,500 addresses (4.5% of the sample) were obtained through application of ACE methods implemented in the field.

Households at the sampled addresses were screened to find members for the shadow sample (ages 9 to 11—identified for use in subsequent Waves but not interviewed in Wave 1), for the youth sample (ages 12 to 17—to participate in the youth cohort), and for the adult sample (partitioned into eight strata based on the cross-classification of age [18-24, 25+], race [Black, other], and tobacco use status [user, non-user]).

Two screeners were used, the first with a household respondent to obtain needed data about all household members, and the second with each sampled adult to obtain self-reported information, in particular that regarding tobacco use. Virtually all age-eligible youth identified through the household screening process were selected to be in the youth sample (for households with more than two youths, a maximum of two youths were selected with the exception of households with multiple-birth youths from which a maximum of three youths were selected—thus, subsampling was employed in most households with three or more youths). Additional subsampling of adults initially selected for the second screener was applied in some strata based on their self-reported data to that screener. Once selected by this process, all people (youth and adults) responding to the Wave 1 interview were considered enrolled in the Wave 1 PATH Study cohort sample and thus eligible to be interviewed in subsequent waves.

Youth in the Wave 1 shadow sample who are permitted by a parent or guardian to participate in the study and have reached age 12 at Wave 4 will be interviewed for the first time. Similarly, 15 year olds in the youth sample at Wave 1 who reach age 18 at Wave 4 will be interviewed using the adult instrument.

### ***Summary of Key Features of the Sample Design for the Replenishment Sample Selected for Wave 4***

The Wave 4 sample design has two important objectives: supplementation of the sample respondents selected at Wave 1, and selection of a shadow sample for fielding in Wave 5. These will be accomplished as follows:

- Sample within the existing PATH Study sample segments from among the addresses not previously selected. (Analysis has indicated that sufficient numbers of addresses are available within the originally sampled segments.)
- Select a sufficiently large sample of addresses to achieve desired sample sizes—note that meeting the sample size needs for the Wave 4 shadow sample (to satisfy Wave 5 youth sample size requirements for 12 and 13 year olds) will produce more than enough addresses for replenishment needs.
- For a randomly selected subset of the sampled addresses (close to two-thirds of the addresses), screen solely for the purpose of identifying members of the shadow sample.
- For the remaining (complementary) subset of addresses, screen for both the shadow sample and the replenishment sample needed for the youth and adult interviews.
- Screening and sampling within households will be carried out as was done for Wave 1, although the sampling rates will vary depending on the replenishment needs of the various sample domains.

### ***Additional Details on the Design of the Wave 4 Replenishment Sample***

Within the PSUs and segments sampled for the PATH Study, addresses will be selected using ABS procedures described earlier. The study will sample close to 180,000 addresses (179,453) for fielding purposes with the expectation that, after accounting for nonresponse (55% screener response rate) and vacancies (12.3% of sampled addresses are assumed to contain no

household eligible for sample selection), the sampling will yield completed household screeners from roughly 87,000 households (86,559).

The subgroup that drives the screening sample size is the 10 and 11 year olds required for the shadow sample. This group requires no “content” data as they will not be asked to complete an interview until Wave 5. The group requiring the largest amount of screening to fill targeted sample sizes and for which interviews will be conducted is “tobacco users who are Black and 18 to 24 years of age.” Based on targeted sample size, relative rarity in the general population, and expected level of misclassification of tobacco use status by the household respondent (based on the Wave 1 experience), this group will require about 30,000 completed household screeners (29,900). The estimate of about 30,000 represents about 34.5 percent of the 87,000 household screeners targeted for completion overall. The remaining sampled addresses will be focused strictly on screening households to identify those containing 10 or 11 year olds who will participate in the study in Wave 5. Thus, the 180,000 sampled addresses will be partitioned into two groups: one of roughly 62,000 (61,988) and the other of about 117,000 (117,465).

As discussed in section B.1.a, a new cohort (Wave 4 cohort) will be formed when the replenishment sample selected at Wave 4 is pooled with the portion of the Wave 1 cohort who are also members of the civilian, household population at the time of Wave 4. Tables B-1 and B-2 show that the expected number of Wave 4 respondents (continuing sample plus replenishment sample) exceeds that for Wave 1 for the domains of interest associated with youth and young adults (18 to 24) and equals that for Wave 1 for the domains associated with older adults. An aim for the Wave 4 sample design was to achieve sample sizes for these domains that are at least as large as those achieved in Wave 1.

However, increases in design effects will be incurred in Wave 4 cohort estimates compared to Wave 1. This is due to variation in sample weights associated with nonresponse adjustments to the weights as well as the aging up of youth to young adults and young adults to older adults, and the pooling of subgroups with different initial sampling rates. To help compensate for these design effects, the sample sizes for youth and young adults have been

targeted somewhat higher than those achieved in Wave 1. For adults 25 or older, simply matching the Wave 4 sample sizes to those for Wave 1 will result in reductions in effective sample sizes due to the increase in design effects. However, the “25+” domains have the largest sample sizes, so issues of precision and power are of less concern. Moreover, over time, the sample size of the “25+” domains will be enhanced with the aging up of youth and young adults to the “25+” categories. One uncertainty related to precision is the extent to which those 18 or older will transition from “user” status at the time of sample selection to that of “non-user” at a later point in time, or vice versa. Such movement will add to the variation of study estimates.

There will be less variation in weights for those under the age of 18 compared to adults because sampling rates for youth are unaffected by user status and race. The number of shadow sample respondents expected in Wave 4 is 4,684, which is expected to provide about 4,200 Wave 5 respondents in the 12 to 13 year old age range. For Wave 5, the projected number of 14 to 15 year olds is about 4,700, while that of 16 to 17 year olds is a little less than 4,600.

### ***Note on Biospecimen Samples***

The plans for biospecimen collection at Wave 4 for continuing adults are described in Section B.2d. In summary, urine and blood samples will be requested from new adults and persons who have aged up to the adult cohort at Wave 4. In addition, urine samples will be requested from all youth participants.

### **B.1d Estimated Response Rates**

For the continuing sample (members of the Wave 1 cohort—those selected at Wave 1 including the shadow sample), two types of response rates are computed. One rate is the retention rate, reflecting the extent to which respondents in the previous wave completed the same instrument (youth or

adult) in the current wave. The other rate is the recruitment rate, reflecting the extent to which sampled persons (shadow or youth) completed a new instrument (youth or adult, respectively).

For the replenishment sample (those to be pooled with the civilian, household respondents among the continuing sample at the time of Wave 4), response rates are associated with the household screener and the extended interview (youth or adult).

### ***Continuing Sample (Wave 1 Cohort)***

For Wave 1, the unweighted response rates for the household screener and extended interviews were 54 percent for households, 75 percent for sampled adults, and 78 percent for sampled youth. For Wave 2, the retention rates were 83 percent for continuing adults and 88 percent for continuing youth. The recruitment rates were 86 percent for aged-up adults (i.e., youth in Wave 1 who reached age 18 at Wave 2) and 82 percent for aged-up youth (i.e., shadow youth in Wave 1 who reached age 12 at Wave 2).

The projected retention and recruitment rates for Wave 3 reflect the experience from Waves 1 and 2 as well as the information from the Medical Expenditure Panel Surveys (MEPS) and the National Longitudinal Survey of Youth (NLSY).<sup>6</sup> Among the Wave 2 respondents, the projected retention rates

---

<sup>6</sup> Few recent in-person longitudinal surveys in the U.S. are directly comparable to the PATH Study. MEPS, like the PATH Study, conducts in-person interviews with respondents, and provides recent data on retention of adults in a longitudinal study, where the retention rate is the complement of the attrition rate. 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 percent and year-2 retention at 95 percent. The retention rates for more recent years of MEPS have been in line with these published rates. Conservative values are used 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 2012-2013 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

for Wave 3 are 85 percent for continuing adults and 89 percent for continuing youth; the projected recruitment rates are 89 percent for aged-up adults and 90 percent for aged-up youth. To help boost the Wave 3 sample yield, most Wave 2 nonrespondents are fielded in Wave 3 (doing so means that one cannot calculate PATH Study overall response rates as the product of successive response rates since Wave 1).

For Wave 4, the projected retention rates among the Wave 3 respondents are 89 percent for continuing adults and 91 percent for continuing youth. The estimated recruitment rates among the Wave 3 respondents are 89 percent for aged-up adults and 90 percent for aged-up youth. Similar to what was done in Wave 3, the study plans to increase the sample yield for Wave 4 by converting Wave 1 respondents who were nonrespondents at Wave 2 and/or Wave 3.

### ***Replenishment Sample***

Wave 4 includes the continuing sample (i.e., the cohort established at Wave 1) and a replenishment sample to bring sample sizes up to at least the Wave 1 sample sizes in terms of numbers of respondents. Projected response rates for the replenishment sample at Wave 4 correspond to those at Wave 1, namely, the response rates of approximately 55 percent at the household screener phase, approximately 75 percent response to the adult extended interview, and approximately 78 percent for the youth extended interview.

---

rate of 86.7 percent (National Institutes of Health, 2010). The projected retention rates for the PATH Study are less than or equal to the 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%, 91%, and 96%, 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.

## **B.2 Procedures for the Collection of Information**

This section includes a brief overview and description of the PATH Study data and biospecimen collection plans for Wave 4. It also discusses the study efforts to minimize duplication with other data collections and the burden such collections place on participants and the public, weighting and estimation procedures, and expected levels of precision.

### **B.2a Overview**

In addition to cohort replenishment at Wave 4, the PATH Study will collect data and biospecimens from new, continuing, and aged-up participants. The study will conduct two parallel screening efforts in the original PSUs and segments to replenish the sample: (1) in-person screening of addresses and recruitment of eligible new adults, youth, and shadow youth; and (2) mail and in-person screening of addresses and recruitment of eligible new shadow youth only.

Wave 4 will include all Wave 3 respondents and nonrespondents from Waves 2 and 3 except: SPs who specifically requested withdrawal from the study; SPs who refused in both Wave 2 and Wave 3 or were firm or hostile refusers (as defined in Section B.1c); SPs who were unable to complete a Wave 3 interview in English or Spanish; SPs with a physical or mental disability or chronic illness that prevents their participation; and deceased SPs.

Wave 4 data and biospecimen collection involves five components: (1) automated computer-assisted personal interviewing (CAPI) household screener instrument for sampled addresses designated as “replenishment;” (2) mail screener instrument for sampled addresses designated as “shadow youth only” with eligible responding households also receiving the CAPI household screener instrument; (3) automated audio computer-assisted self-interviewing (ACASI) instruments, with separate instruments for youth and adults; (4) CAPI parent instrument; and (5) collection of biospecimens from adults and youth. Collection of biospecimens is not a requirement for

participation; however, completion of an extended interview is required for biospecimen collection.

Data collection components and instruments in Wave 4 will also differ for: (1) newly selected households at replenishment sample addresses, (2) newly selected households at shadow youth only addresses, (3) continuing adult SPs, (4) continuing youth SPs and their parents, (5) new adults and youth SPs who age up to the adult cohort, and (6) new youth and shadow youth who age up to the youth cohort and their parents.

The PATH Study field interviewers are responsible for obtaining complete and accurate information from the addresses and SPs assigned to them. Field interviewers receive extensive training on procedures for administration of data collection instruments, including techniques to establish rapport and gain cooperation, to explain the importance of the PATH Study to the respondent, and to answer respondent questions or address any concerns; and on procedures for the collection of biospecimens. All field interviewers will receive a the Wave 4 Field Procedures Manual, which details every procedure required for their work.

In preparation for the start of Wave 4, the PATH Study requires field interviewers to complete home study, in-person, and web-based trainings. Newly hired field interviewers participate in a 16-hour home study training that focuses on using the laptop computer and other equipment, study materials, making in-person contact with residents at replenishment sample addresses, making initial telephone contact with continuing respondents, administrative tasks, and managing workloads and assignments. Any newly hired field interviewers without recent experience in household data collection receive an additional six hours of home study training in general interviewing techniques and standards for interviewer behavior. After the home study training, newly hired interviewers participate in six days of in-person, in-depth training on data and biospecimen collection procedures, including: (1) conducting CAPI screening interviews at replenishment sample addresses; (2) verifying by phone or in person each continuing SP's contact information and scheduling in-person interviews; (3) techniques for obtaining consent; (4) conducting the ACASI interviews and CAPI parent interview; (5)

collecting UPC codes on tobacco products with the UPC code scanner; (6) collecting and shipping urine samples; (7) scheduling blood collections; (8) issuing respondent incentives; and (9) completing administrative procedures, such as data transmission and reporting to a field supervisor. Throughout the training, interviewers practice using the PATH Study laptop computer and other equipment.

Experienced field interviewers from Wave 3 will participate in an eight-hour home study that covers the Wave 4 protocol as well as in three days of in-person, in-depth training on new tasks for Wave 4. Experienced phlebotomists will receive training on PATH Study procedures for visiting the homes of consenting adults to collect blood samples; all phlebotomists will receive the PATH Study's Phlebotomist Manual on the study's protocol for blood collection.

All in-person trainings include certification for field work. In this step, trainees are observed by training staff as the trainee performs all procedures that may be part of a data collection visit. For those who are not proficient in all tasks, additional practice and an opportunity for recertification is offered. No interviewer or phlebotomist begins work on the PATH Study until his or her supervisor certifies performance aligns with the study protocol and procedures.

Quality control procedures used by the PATH Study ensure that field interviewers are following specified protocol and procedures. These include review of audio recordings of interviews, in-person or telephone validation, and review of global positioning system (GPS) data recorded on interviewer laptops during data collection. Field interviewers who show any potential weakness at any point in the data collection are observed in-person at least one time by a field supervisor or quality control interviewer. 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.

As part of quality control, the PATH Study validates 10 percent of data collection by each interviewer throughout the wave. Validation may be by means of the audio recordings of selected items from the CAPI portions of data collection with adults. These recordings, with the consent of respondents, use computer-assisted recorded interviewing. If a respondent refuses audio-recording or as otherwise needed, quality control validation interviews are conducted by telephone or in person by trained staff. For some non-complete dispositions (e.g., unable to make contact with an SP), trained staff validate the disposition in person. Validation confirms that an interview was administered or attempted as reported by the field interviewer. (See Attach2.PATH Study Data Collection Instruments for Validation Forms.) The PATH Study validates 100 percent of an interviewer's data collection if any concerns arise about his/her performance.

Another quality control tool is the use of GPS data recorded on field interviewers' laptops during data collection. The PATH Study uses these data, with case information, to check that the address location at the time of data collection matches the case address. Cases that fail such data quality checks are reported to field management staff for additional review and follow up. Any suspect findings are fully investigated and resolved.

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. As needed, based on feedback on field interviewer performance, field management staff retrain field interviewers.

Management staff located at both the home office and remote sites have access to a web-based 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 reports include weekly reports on progress during the past week as well as on potentially suspicious field interviewer behavior, such as 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 Screener Interview**

The Wave 4 screening of households for the PATH Study replenishment sample will be conducted through in-person screening for new adults, new youth, and new shadow youth; and mail and in-person screening for only new shadow youth. The in-person screening for new adults, youths, and shadow youths involves the random selection of up to two adults and two youths (unless a household includes multiple-birth youths, in which case additional youths could be selected) per eligible household (as described in Section B.1) and is conducted using an automated sampling algorithm programmed within the CAPI screening instrument (see Attach2.PATH Study Data Collection Instruments). The screener respondent will be an adult household member age 18 or older. The in-person screener uses a full household enumeration process to collect information on the age of each household member as well as the race, active military service status, ability to speak in English or Spanish, and tobacco use status of each adult household member; the screener also collects information on how the screener respondent and household members are related. To allow the recontact of the household for quality control purposes, or to set appointments for the extended and parent interviews if the SP is unavailable at the time of the screening, contact information (telephone numbers and email addresses) is collected for the screener respondent, each adult SP, and the identified parent of each youth SP in the household. Finally, if the mailing address differs from the street address, the household mailing address is collected. The mailing address allows written follow-up with nonresponse cases and regular contact with respondents between data and biospecimen collection waves, as discussed in Section B.3.

To screen for additional new shadow youth, the study will initially mail a hardcopy form to a subset of randomly selected addresses. If a household does not respond to the initial and follow-up mailings, a field interviewer will visit the address and administer the CAPI household screener described in

the previous paragraph, to enumerate the household and select the shadow sample youth. If a household does respond to a mailing and reports having at least one youth aged 10 or 11 living in the household, the home office will mail an advance letter to the household, and a field interviewer will visit the address and administer the CAPI household screener.

### **B.2c Extended Interview**

Wave 4 data collection procedures differ for: (1) continuing adult SPs, (2) continuing youth SPs and their parents, (3) new adults and youth SPs who age up to the adult cohort, and (4) new youth and shadow youth who age up to the youth cohort and their parents. Approximately 3 months in advance of the anniversary of the earliest Wave 3 interview conducted for a continuing SP's household (household anniversary month),<sup>7</sup> the home office will send a letter to adult SPs and parents of youth SPs to remind them of the upcoming follow-up interview. Approximately 1 month in advance of the household anniversary month, the field interviewer will contact an adult in the household by telephone to update contact information on study participants and arrange a convenient time for the in-person data collection visit(s) at the SP's home. (See Attach2.PATH Study Data Collection Instruments for Verification Form) At the time of the in-person household screening, the field interviewer will seek to collect data and biospecimens from new adults and new youth and their parents; or schedule an appointment for a return visit or, at minimum, determine the best time for a return visit.

#### ***Continuing Adult SPs***

For adult SPs from a previous wave, at the time of the in-person visit for Wave 4, the field interviewer will: (1) review the main elements of the informed consent obtained at a previous wave;(2) administer the Wave 4 adult extended interview, including updating contact information; (3) review

---

<sup>7</sup> To increase operational efficiency in Wave 4, all adult SPs and parents of youth SPs in the same household will be contacted on the same schedule. This schedule is based on the anniversary of the earliest Wave 3 interview conducted with an SP in a household.

the main elements of consent for the biospecimen collection and request a urine sample from a subsample of continuing adults who provided urine at a previous wave and collect the urine sample; and (4) pay the incentive to the respondent. (The biospecimen collection is discussed further in Section B.2d.) If an adult SP is unavailable or unable to complete the interview at the scheduled time, the field interviewer will attempt 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 will launch the ACASI extended interview, which begins with an optional tutorial on using ACASI. As required throughout the interview, the field interviewer will remain available to aid the SP in use of ACASI and to respond to questions the SP may have. At the end of the ACASI extended interview, the field interviewer will update the SP's contact information.

The adult SP who completes the extended interview will receive \$35 (the adult extended interview incentive) on a PATH Study debit card as a thank you and a thank you card (Attach9.Follow-up, Retention, and Tracking Materials). A refusal conversion letter will be sent to adult SPs who initially decline to participate or are difficult to contact (Attach18.Field Data Collection Materials). 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.

### ***Continuing Youth SPs***

For youth SPs from a previous wave, the field interviewer will: (1) review with the parent the main elements of parental permission for the youth's participation obtained at a previous wave; (2) request parental permission for youth urine collection; (3) review with the parent the main elements of consent for the parent interview; and (4) administer the CAPI parent

interview, which includes updating the parent’s contact information.<sup>8</sup> Field interviewers cannot conduct youth interviews before completing review of parental permission with the parent; also, field interviewers cannot collect urine samples from youth before receiving parental permission for that collection. The parent who completes a parent interview for the youth will receive \$10 on a PATH Study debit card as a thank you and a thank you card.

If a youth SP from a previous wave with parental permission is available and has time at the visit to complete the interview, the field interviewer will: (1) review the main elements of assent with the youth SP, obtained at a previous wave; (2) administer the ACASI extended interview; (3) after the extended interview, request youth assent for urine collection; (4) if the youth assents, collect a urine sample; and (5) pay the incentive to the respondent. (The biospecimen collection is discussed further in Section B.2d.) The youth SP who completes the extended interview will receive \$25 (the youth extended interview incentive) on a PATH Study debit card and a certificate of appreciation that acknowledges the youth’s contributions/community service as a participant in the PATH Study (Attach9.Follow-up, Retention, and Tracking Materials). A refusal conversion letter will be sent to parents who initially refuse or are difficult to contact (Attach18.Field Data Collection Materials). A youth SP may also receive \$5 on up to two occasions when his/her parent updates the parent’s contact information during the year, for a total of \$10.

### ***New Adults and Youth SPs Who Age up to the Adult Cohort***

At the in-person visit, the field interviewer will: (1) attempt to obtain informed consent (Attach11.Consent Materials); (2) administer the adult extended interview, including gathering additional contact information about the adult; (3) request consent for the biospecimen collections; (4) if the adult consents, collect the urine sample; (5) arrange a follow-up appointment for a

---

<sup>8</sup> At Wave 4, if the youth SP is living with a parent other than the one who originally permitted youth participation, the field interviewer requests permission and consent from the “new” parent and conducts the parent interview with him/her.

phlebotomist to collect a blood sample; and (6) pay the incentive to the respondent at the completion of the first home visit. (The biospecimen collection is discussed further in Section B.2d.) For new adults, the first part of the extended interview is the individual screener. Responses to the items in this screener may confirm or contradict the information provided in the first-phase household screener by the household screener respondent. Depending on the individual's self-reports (e.g., on tobacco use), a new adult may be de-selected and not asked to complete the remainder of the extended instrument; all aged-up adults will receive the full extended interview.

### ***New Youth and Shadow Youth Who Age up to the Youth Cohort***

For the selected new youth and youth who has aged up to the youth cohort, following parental permission, the interviewer will: (1) attempt to obtain youth assent (Attach11.Consent Materials); (2) administer the automated ACASI extended youth instrument; (3) after the extended interview, request youth assent for urine collection; and (4) if the youth assents, collect a urine sample. (The biospecimen collection is discussed further in Section B.2d.)

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

Wave 4 interviews for adults and youth are designed to build on information collected at previous waves. Thus, relatively stable information such as demographic characteristics (e.g., sex and race) is collected only at the respondent's "baseline" or first wave. Similarly, information on lifetime use of tobacco products is not asked again for products a respondent reported having used at his or her baseline wave. Not repeating questions from the previous interview about established characteristics or past behaviors will help keep respondent burden at Wave 4 to an average of 1 hour for the adult interview and 35 minutes for the youth interview.

The parent interview collects information about the parent of a sampled youth, the household, and the youth, as well as contact information to reach the parent 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 for more than one youth. In such cases, the parent will not be asked to provide the same information again, but only information relevant to the particular youth.

**B.2d Biospecimens**

As noted in Section B.1d, field interviewers will request urine samples at Wave 4 from a subsample continuing adults who initially provided urine at a previous wave, with the expectation of collecting samples from approximately 11,000 of these adults. (See Table B-3.) They will also collect urine and arrange for collection of blood samples from new adults and youth SPs who age up to the adult cohort at Wave 4 and consent to biospecimen collection. Also, field interviewers will collect urine samples from assenting youth with parental permission.

Although completion of the extended interview is required from all respondents to participate in the longitudinal cohort, providing biospecimens is voluntary and not a condition of participation. Respondents will receive \$25 on a PATH Study debit card for participating in the urine sample component of the study and \$25 for participating in the blood sample component.

Table B-3. Summary of plans for biospecimen collection at Wave 4

Type of respondent	Biospecimen	
	Urine	Blood
Continuing adults who initially provided urine at a previous wave	Yes, from a subsample of approximately 11,000 adults	No
New adults and aged-up adults	Yes	Yes
Continuing youth, new youth, and aged-up youth	Yes	No

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

Sample weights are developed for the PATH Study respondents at each wave to permit estimation for and inference about the target populations from which the sample was drawn. As discussed in section B.1a for Wave 4, the study has two target populations of interest and hence two sets of weights are needed.

One of the two target populations comprises members of the U.S. civilian household population at the time of Wave 1 who are living in the U.S. in Wave 4 (including Wave 1 respondents now living in institutions or members of the military). This population is represented by the continuing sample of the Wave 4 respondents first sampled for Wave 1; it will be referred to as the Wave 1 cohort. The Wave 1 cohort can serve as the basis for longitudinal analyses of the data collected up through Wave 4. As later waves of data become available, the continuation of this cohort will provide the data for extended longitudinal analyses.

The other target population comprises the U.S. civilian household population at the time of Wave 4. This population is represented by the combination of the replenishment sample and the Wave 4 respondents from the Wave 1 cohort who, in Wave 4, are members of the civilian, household population; it will be referred to as the Wave 4 cohort. The Wave 4 cohort can serve as the basis for cross-sectional estimates for the civilian, household population at the time of Wave 4 as well as the starting point for longitudinal analyses of this cohort when later waves of data become available.

The main objectives associated with the development of both sets of sample weights for Wave 4 are to:

- Permit the appropriate development of estimates, reflecting the probabilities of selection of a respondent, which can vary depending on factors such as the cohort of interest and residential status in Wave 1;

- Limit the potential for biases arising from differences between cooperating and non-cooperating SPs and households;
- Limit 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.

Data used for weighting will be edited, checking for unexpected values or combinations of values and expected consistencies as well as examining frequency distributions. The developments of the two sets of weights are described separately below.

### ***Developing Weights for Members of the Wave 1 Cohort***

The sample design for the Wave 1 cohort of the PATH Study involved the sample selection of members of the U.S. civilian, household population at the time of Wave 1. The sampled persons are then followed as long as they remain in the U.S. resident population, and the 12 year old population has been replenished annually from a shadow sample of 9 to 11 year olds also sampled during Wave 1. Thus, for Wave 4, the sample weights of respondents represent, in total, the resident population of the U.S. ages 12 or older at the time of Wave 4 who were in the civilian household population during Wave 1.

All Wave 1 cohort longitudinal weights are based on the original Wave 1 cross-sectional weights previously established. The process for computing these weights was described in detail in Section B.2e of Supporting Statement B for Wave 1. The basic steps were:

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, and “raking” the nonresponse-adjusted weights to population control totals;

3. Creating person base weights by multiplying the household nonresponse-adjusted weights by the reciprocal of the selection probabilities of SPs;
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 balanced repeated replication method for use in variance estimation.

After Wave 1, the development of the longitudinal weights for each subsequent Wave J follows the two steps below:

1. Adjust for the nonresponse incurred since Wave 1.
2. Rake the resulting nonresponse adjusted weights to sample-based control totals (discussed below) with trimming incorporated as part of the raking process, as needed.

One set of Wave 1 weights was created for all youth who completed the Wave 1 interview and another set was created for all adults who completed the Wave 1 interview. Weights were also created for the 9 to 11 year olds selected as part of the shadow sample in Wave 1 for use in later waves as the “base weights” for the shadow sample members when they become 12 year olds and are asked to complete the youth interview.

The final Wave 1 adult, youth, and shadow youth weights will serve as the starting point (“base weights”) for the development of weights for Wave 4.<sup>9</sup> These Wave 1 weights sum to the sizes of the eligible populations of adults, youth, and shadow youth, respectively, in 2013 (since the final step of the Wave 1 weighting involved raking to 2013 ACS control totals). Some of the

---

<sup>9</sup> Wave 1 weights, rather than Wave 3 weights, are used as the starting point because, in addition to Wave 1, some persons will complete interviews in Wave 4 but perhaps not in Wave 3 and/or Wave 2.

Wave 1 respondents will have died or left the U.S. at the time of Wave 4. The target population for the Wave 1 cohort at the time of Wave 4 is thus those members of the Wave 1 target population living in the U.S. at the time of Wave 4. As a result, the sum of the Wave 4 longitudinal weights will be slightly smaller than the sum of the Wave 1 weights.

The first step in creating Wave 4 weights for this cohort will be to adjust the “base weights” of the Wave 4 respondents to account for attrition between Wave 1 and Wave 4, following a standard approach used to compensate for wave nonresponse (see, for example, Kalton, 1986; and Brick, 2013). Särndal and Swensson (1987) discuss approaching nonresponse adjustments as analogous to two-phase sampling, which would allow use of Wave 1 (and possibly Wave 2 or 3) interview data from all Wave 1 adults and youth to be used to construct weights for continuing adults and youth who participate at Wave 4. Nonresponse adjustment cells will be formed using available variables including age, race, ethnicity, sex, employment status, education level, tobacco use, household composition, and census block characteristics. For Wave 1 shadow youth who have their first youth interview at Wave 4 (9 year olds at Wave 1), information from the Wave 1 household screener will be used to form nonresponse adjustment cells; for Wave 1 shadow youth who were interviewed as youth at Wave 2 or 3, the data collected in the corresponding wave can be used to form the Wave 4 nonresponse adjustment cells.

Weight adjustments will be computed within cells formed from the cross-classification of available variables. The tree-based classification software CHAID will be employed to identify cells that distinguish between subgroups with different propensities to respond to the PATH Study over time (see Roth et al., 2006; and Schouten and deNoij, 2005 for discussions of tree-based classification). Then weighting adjustment factors will be determined, after accounting for standard issues such as small cell sizes and large adjustment factors.

Raking to sample-based control totals can help limit drifting from some important baseline characteristics that might arise through the applications of nonresponse adjustments over time. Lundstrom and Sarndal (1999)

provide a theoretical discussion of the use of calibrating weights to sample-based controls as well as providing empirical evidence that such calibration can serve to reduce both variance and nonresponse bias.

This raking will be accomplished for the Wave 4 respondents who are members of the Wave 1 cohort by establishing a separate set of raking dimensions for each replicate weight, where the dimensions that are population-based will be a constant across all replicate weights while those that are sample-based will vary based on the application of the replicate weight to Wave 1 data. In this fashion, the variation associated with the Wave 1 PATH Study estimates serving as sample-based controls will be appropriately reflected in the variance estimates for Wave 1 cohort estimates, with the expectation of reducing both overall variance and the potential for bias.

### ***Developing Weights for Members of the Wave 4 Cohort***

Most members of the Wave 1 cohort will also be members of the Wave 4 cohort, and most individuals selected for the replenishment sample at Wave 4 will have been eligible to have been selected for the Wave 1 sample. Thus, in order to fully reflect the civilian, household population in Wave 4, care will be taken to deal with issues of double chances of selection as well as ensuring that only members of the Wave 4 cohort target population are included in the weighting process.

Interview questions will be asked of the respondents in the Wave 4 replenishment sample to determine if they were members of the U.S. civilian, household sample at the time of Wave 1. If not, they have only a single chance of selection (through the replenishment sample) and their weights will be determined accordingly. All other members of the Wave 4 cohort will have two chances of selection, either through the Wave 1 sample or the replenishment sample. To account for this, the weights of the two components of the sample with two chances of selection will be composited as discussed below. Note that Wave 4 continuing sample respondents who are in the military or in an institution will not be included in the cross-

sectional weighting as they are not members of the civilian, household population at the time of Wave 4. Thus, they will not be members of the Wave 4 cohort for analysis purposes.

Development of the Wave 4 cohort weights will follow the Wave 1 weighting process outlined above through the computation of person base weights (Step 3) and nonresponse adjustments (Step 4). The adjustments will be undertaken separately within the Wave 1 cohort sample and the replenishment sample. For persons with two chances of selection, compositing factors will be applied, adjusting the weights of those from the continuing sample by a factor  $\alpha \leq 1$  and those from the replenishment sample by a factor  $(1 - \alpha)$ , in compensation. Different values of  $\alpha$  will be applied for the respondents within each of the various subgroups associated with sampling for the PATH Study (e.g., Black adults ages 18 to 24 who are tobacco users). Within each such subgroup, the effective sample sizes (i.e., the sample sizes divided by the design effects associated with the variation of the weights) for the two sample components will be used to determine the value of  $\alpha$  for the subgroup. The full set of weights will be raked to ACS control totals. Replicate weights will be created for variance estimation purposes using balanced repeated replication methodology. The capability for undertaking variance estimation using Taylor Series linearization will also be established.

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

The PATH Study is designed to produce reliable estimates of differences between population subgroups and over time, including within-person changes in tobacco-related attitudes, behaviors, and health conditions. Many characteristics of interest are dichotomous, having “yes” or “no” outcomes. The percentage of “yes” responses can be denoted by  $p$ , representing the prevalence estimate for a particular characteristic (e.g., cigarette smoking). Past and current research suggests that most of the characteristics measured in the PATH Study will have magnitudes of prevalence that exceed 10 percent. However, a few, such as initiation of tobacco use, can be expected to fall between 1 and 5 percent.

One measure of the precision associated with cross-sectional prevalence rates is the relative standard error (RSE), defined as the standard error of an estimate 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). Here, the MDADs represent the smallest change (up or down) from a given Wave 1 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 PATH 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 design.

Clustering at both the PSU and segment levels in the PATH Study's sample design contributes to its overall design effect. For a fixed sample size, the greater the number of responding units per cluster and/or the homogeneity of the respondents with respect to a characteristic of interest within clusters, the greater the DEFF and, hence, inflation of the variance (resulting in decreased precision). The level of homogeneity within clusters 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 actually strata, not PSUs;

therefore, these PSUs do not contribute to the variance from clustering at the PSU level. Thirty-five of the 156 PSUs selected are certainties, representing 24 percent of the U.S. population.

Another feature is the sampling of adults with different selection probabilities according to their age, race, and tobacco use (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 can range substantially, depending on the demographic or tobacco use domain of interest. For analyses that combine all adult respondents, this component of the unequal weighting DEFF has been estimated to be approximately 2.1.

A third feature is the restriction that no more than two adults can 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 use 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 (RSEs) and power (MDADs) for the Wave 1 cohort at Wave 4 are calculated after taking into account the DEFFs resulting from the three previously-described sample design features. These estimates are shown in Tables B-4 and B-5, for adults and youth, respectively. Only RSEs are provided in Tables B-6 and B-7 for the Wave 4 cohort because it represents the initial wave associated with this cohort. The projected RSEs are for a generic statistic estimating a prevalence rate of 15 percent (e.g., hypothetically, the percentage of the adult population who are every day cigarette smokers). The MDADs presented are for a generic statistic estimating change from a Wave 1 prevalence rate of 10 percent to a

corresponding estimate in Wave 4. Both the RSEs and MDADs presented here are for illustrative purposes.

**Estimates Based on Wave 1 Cohort Data**

In Tables B-4 and B-5, the RSEs are for Wave 4 estimates based solely on Wave 1 cohort data (i.e., the continuing sample). The MDADs are for a change from Wave 1 to Wave 4. 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 age into the older group of adults. As a result, variation in weights among members of most subgroups will increase with a corresponding increase in DEFF values due to unequal weighting. Inflation factors were computed separately for adults 18 to 24 years old (where at Wave 4, the persons ages 18 to 20 were originally sampled as youth) and for adults ages 25 and over. For adults 18 to 24 years old, the inflation factors range from 1.01 to 1.32 and for adults ages 25 and over, the inflation factors range from 1.01 to 1.06. The estimates of 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 should be interpreted with caution.

Table B-4. Adult sample sizes, relative standard errors (RSEs), and minimum detectable absolute differences (MDADs) for Wave 4 estimates based on **Wave 1 cohort** data

<b>Group</b>	<b>Wave 1 cohort sample size at Wave 4</b>	<b>RSE on 15% item</b>	<b>MDAD on 10% item</b>
<b>All adults</b>	<b>24,998</b>	<b>2.9%</b>	<b>0.8%</b>
Wide-net users	16,573	2.9	0.8
Current and experimental users	13,500	3.0	0.9
Menthol cigarette smokers	3,393	5.0	1.3

<b>Group</b>	<b>Wave 1 cohort sample size at Wave 4</b>	<b>RSE on 15% item</b>	<b>MDAD on 10% item</b>
Dual (smokers and smokeless tobacco users)	801	9.6	2.5
Daily users	8,153	3.6	1.0
Less-than-daily users	3,913	4.7	1.2
Current non-users under wide-net definition	8,425	3.9	1.1
<b>Adults ages 18-24</b>	<b>7,888</b>	<b>3.7</b>	<b>1.5</b>

Table B-5. Youth sample sizes, relative standard errors (RSEs), and minimum detectable absolute differences (MDADs) for Wave 4 estimates based on **Wave 1 cohort** data

<b>Group</b>	<b>Wave 1 cohort sample size at Wave 4</b>	<b>RSE on 15% item</b>	<b>MDAD on 10% item</b>
<b>All youth</b>	<b>10,365</b>	<b>3.2%</b>	<b>1.2%</b>
Current tobacco users	871	8.8	3.2
Current cigarette smokers	481	11.7	4.2
Menthol cigarette smokers	291	14.9	5.4
Experimenters	1,141	7.7	2.8
Never cigarette smokers	8,972	3.3	1.3
Susceptible never cigarette smokers	1,865	6.2	2.3
Never tobacco users	5,940	3.8	1.5
<b>Youth ages 12 to 13</b>	<b>3,588</b>	<b>4.7</b>	<b>2.2</b>
Current tobacco users	61	32.4	15.1
Current cigarette smokers	37	41.5	19.3
Menthol cigarette smokers	20	56.4	26.2
Experimenters	157	20.2	9.4
Never cigarette smokers	3,423	4.7	2.3
Susceptible never cigarette smokers	491	11.5	5.4
Never tobacco users	2,239	5.7	2.7
<b>Youth ages 14 to 17</b>	<b>6,777</b>	<b>3.7</b>	<b>1.6</b>
Current tobacco users	811	9.1	3.8
Current cigarette smokers	444	12.1	5.0
Menthol cigarette smokers	271	15.4	6.4
Experimenters	985	8.3	3.4
Never cigarette smokers	5,549	3.9	1.7
Susceptible never cigarette smokers	1,374	7.1	3.0
Never tobacco users	3,700	4.6	2.0

The sample of adult tobacco users at Wave 4 for the Wave 1 cohort will be sufficiently large to allow analyses for many user subgroups as well as for persons who are considered at risk of becoming tobacco users. Table B-4 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 year olds). The subgroup sample sizes for the different categories of tobacco users were estimated using data from the full Wave 1 sample of the PATH Study. A “current user” definition was applied in estimating sample sizes for menthol cigarette smoker, dual and daily user, and less-than-daily user groups; the “wide net”

definition, provided earlier, was used to estimate the sample sizes for nonuser groups. These definitions give the smallest sample sizes, and hence the largest RSEs and MDADs, for each of these groups. That is, the estimated RSEs and MDADs for another definition of tobacco user will be smaller than those displayed in the tables, so the estimates presented are conservative. The RSEs for a 15 percent prevalence rate are at or below 5 percent for most adult subgroups shown. The MDADs for a 10 percent Wave 1 prevalence rate are mostly in the neighborhood of 1-1.3 percentage points indicating that a three-year change of roughly 1.3 percentage points or less can be reliably detected for the subgroups shown. The adult sample sizes considered in this section are based on estimates for completed Wave 4 interviews; therefore, the estimates of precision and power apply to projected estimates of tobacco and health outcomes collected with the Wave 4 instruments. As is the case for all the estimates presented in this section, the study expects that precision and power will be reduced for finer divisions of the subgroups (e.g., by sex).

The initial sample of 13,651 youth at Wave 1 was intended both to provide sufficient power for analyses of youth subgroups and to enhance the adult cohort in future waves of the PATH Study, as youth age into adulthood. Table B-5 shows expected Wave 1 cohort sample sizes at Wave 4 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 cigarette smokers, susceptible never cigarette 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 the full Wave 1 sample of the PATH Study. For youth, current smokers were defined as youth who have smoked a cigarette within the past 30 days, and current users were youth who have used any tobacco product within the past 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?"

Many of the subgroups are large enough to produce stable estimates. For example, there are approximately 5,940 never users for whom tobacco use initiation rates will be tracked. Tobacco cessation is more of an issue in the older adolescent group (14 to 17 year olds) because more tobacco users are in that age group than among youth ages 12 to 13; the older adolescent group includes about 811 tobacco users and 444 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 cigarette smokers. Subgroups within the PATH Study are generally expected to provide sufficient precision for studying nonusers among the 12 to 13 year olds; however, subgroups with very small sample sizes, such as 12 to 13 year old menthol cigarette smokers, will likely be combined (e.g., 12 to 17 year old menthol cigarette smokers) 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 old ("all youth" in the table) are below 12 percent for all but one of the seven "tobacco use" subgroups and at or below 9 percent for five of them. In terms of MDAD values, among all youth 12 to 17 years old, the sample size in each of the subgroups, except "current" and "menthol" cigarette smokers, is sufficient to detect a three-year change of roughly 3.0 percentage points in a 10 percent Wave 1 behavior overall. Measures of quitting, initiation, and non-cigarette tobacco use tend to be in this 10 percent range (depending on the definitions used).

### ***Estimates Based on Wave 4 Cohort Data***

In Tables B-6 and B-7, the sample sizes and RSEs are for Wave 4 subgroup estimates based on Wave 4 cohort data. The Wave 4 cohort subgroup sample sizes exceed the corresponding ones from the Wave 1 cohort appearing in Tables B-4 and B-5, due to the replenishment sampling. The estimated Wave 4 cohort RSEs take into account the sample-design features

of both the initial sample and the replenishment sample. This includes the enhanced sample sizes as well as the increase in variability in the weights due to factors such as varying sampling rates and weight adjustments.

Table B-6. Adult sample sizes, relative standard errors (RSEs) and minimum detectable absolute differences (MDADs) for Wave 4 estimates based on **Wave 4 cohort** data

<b>Group</b>	<b>Wave 4 cohort sample size at Wave 4</b>	<b>RSE on 15% item</b>
<b>All adults</b>	<b>34,150</b>	<b>2.8%</b>
Wide-net users	23,276	2.7
Current and experimental users	18,903	2.8
Menthol cigarette smokers	4,724	4.4
Dual (smokers and smokeless tobacco users)	1,140	8.1
Daily users	11,283	3.2
Less-than-daily users	5,603	4.1
Current non-users under wide-net definition	10,874	3.5
<b>Adults ages 18-24</b>	<b>10,942</b>	<b>3.4</b>

Table B-7. Youth sample sizes, relative standard errors (RSEs) and minimum detectable absolute differences (MDADs) for Wave 4 estimates based on **Wave 4 cohort** data

<b>Group</b>	<b>Wave 4 cohort sample size at Wave 4</b>	<b>RSE on 15% item</b>
<b>All youth</b>	<b>15,058</b>	<b>2.8%</b>
Current tobacco users	1,271	7.0
Current cigarette smokers	701	9.2
Menthol cigarette smokers	424	11.7
Experimenters	1,663	6.2
Never cigarette smokers	13,028	2.9
Susceptible never cigarette smokers	2,713	5.0
Never tobacco users	8,625	3.3
<b>Youth ages 12 to 13</b>	<b>5,163</b>	<b>3.9</b>
Current tobacco users	87	25.6
Current cigarette smokers	53	32.7
Menthol cigarette smokers	29	44.4
Experimenters	225	16.0
Never cigarette smokers	4,926	3.9
Susceptible never cigarette smokers	707	9.2
Never tobacco users	3,222	4.6
<b>Youth ages 14 to 17</b>	<b>9,895</b>	<b>3.1</b>
Current tobacco users	1,184	7.2
Current cigarette smokers	648	9.6
Menthol cigarette smokers	395	12.1
Experimenters	1,438	6.6
Never cigarette smokers	8,102	3.3
Susceptible never cigarette smokers	2,006	5.7
Never tobacco users	5,403	3.8

Tables B-4 and B-6 present two sets of estimates associated with Wave 4. Table B-4 reflects estimates based on Wave 1 cohort data for Wave 4, while Table B-6 presents corresponding estimates based on Wave 4 cohort data. Both show RSEs for a 15 percent prevalence rate among different categories of adult tobacco users. For the seven “tobacco use” subcategories considered within the full grouping of all adults (18 or older), the median RSE in Table B-4 is 3.6 percent, and the median RSE in Table B-6 is 3.5 percent.

Similarly, estimates based on Wave 4 data on youth appear in Tables B-5 and B-7 reflecting Wave 1 cohort and Wave 4 cohort data, respectively. Again, RSEs for a 15 percent prevalence rate among the seven different categories of youth tobacco users are considered. For these categories and within the grouping of “all youth” (ages 12 to 17), the median Wave 1 cohort

data RSE in Table B-5 is 7.7 percent, and the median Wave 4 cohort data RSE in Table B-7 is 6.2 percent.

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

For Wave 4 and for potential follow-up waves thereafter, the PATH Study will continue to maintain contact with respondents and maximize their retention in the study. The methods currently used by the PATH Study include: (1) tracking participants (through requests to update contact information by visiting the study website, calling a toll-free number, or sending updated information via mail) and tracing them, as needed; (2) maintaining a sufficiently large field interviewer workforce located in or 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, telephone, email, and text messaging<sup>10</sup> in advance of in-home data and biospecimen collection visits; (6) continuing to emphasize the importance of biospecimen collections to field interviewers and respondents; (7) as appropriate, interviewing adults who have relocated to group quarters facilities since their initial interviews; and (8) attempting to convert refusals from adults.

OMB's terms of clearance in approving the Wave 3 data collection require NIDA and FDA to report to OMB regarding: "a) the response rates associated with the full baseline wave [and full Wave 2], including screening, interview completion, and bio-specimen response; b) Wave 3 retention, recruitment rates for the "age in to adult" and "age in of shadow" subsamples; c) the results of nonresponse analysis and statistical approach for addressing non-response, as well as implications for the study going forward; and d) the statistical approach to be applied to the bio-specimen data to address potential non-response bias from lower consent and cooperation rates with

---

<sup>10</sup>Email and text messaging will only be used with adults who agree to be contacted this way.

this aspect of the study.” Section B.4 summarizes the results of the PATH Study’s Interim Report on these topics.

### **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 4 to maintain respondent engagement:

- Mail a thank you card after the Wave 4 interview and a birthday card to adult SPs and the parents of youth SPs for the youth’s birthday;
- For each newly selected shadow youth, meet with a parent and explain the PATH Study would like to enroll the youth once he/she turns 12 years old;
- In households with only shadow youth, mail a follow-up letter to the parents of shadow youth at 6 months after the Wave 4 parent contact and telephone the parents at approximately 12 months after that contact;
- At 12 and 15 months after the Wave 4 interview, mail a follow-up letter to the adults and parents of youth that provides an update on study activities and requests updated contact information;
- Visit respondents who have moved up to 100 miles from a study PSU;<sup>11</sup> and
- When feasible, attempt to visit respondents who have moved more than 100 miles from a study PSU within the U.S.

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

---

<sup>11</sup> The percentage of SPs who move annually beyond 100 miles from any of the Study PSUs is estimated to be less than 1 percent.

provides for real-time access in the field and at the home office to the most current information available.

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 the initial interview for tracing references.** At their initial PATH Study interview, 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 who tend to be mobile, respondents are asked for additional information that may help to locate them.
- **Use interim contacts to determine if contact information has changed or if tracing is needed.** Contacts by mail ask respondents to report any contact information changes. The study provides several ways this can be done, including visiting the study website, calling a toll-free number, or sending updated information via mail. The PATH 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.
- **Update contact information as part of data collection.** During contacts with households for each follow-up wave, the field interviewers update contact information for adults as well as for relatives or persons not living in the household who can serve as references on where to locate the respondent.

The PATH Study approach for tracing and locating respondents who may be lost at the Wave 4 follow-up draws on multiple sources of information in cost-effective ways. If current occupants of the respondent’s 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 a PATH Study tracing specialist for the second line of tracing using the following protocols:

- **BeneVerify.** 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 4 Data and Biospecimen Collection**

To minimize attrition and maximize response rates for the Wave 4 data and biospecimen collection (and for each potential follow-up wave thereafter), the PATH Study employs field interviewers and supervisors who live within or in close proximity to the study PSUs. This helps to ensure that field staff are familiar with the communities within which their assigned cases are located. Field interviewers are also trained in effective techniques for developing rapport and gaining respondent cooperation through refusal aversion and conversion.

In addition to the respondent incentives described in Section A.9, the PATH Study uses several tools and approaches to address nonresponse and maximize response rates. These 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 household anniversary month to inform adult SPs and the parents of youth SPs about the planned Wave 4 data collection (Attach9.Field Data Collection Materials); and follow-up letters that are sent to adults and parents of youth at 12 and 15 months after the Wave 4 interview. The letters remind recipients about the PATH Study objectives, how their data will be used, why the study is important, and why the study includes tobacco users and non-users.
- Respondents can easily access information about the PATH Study through the PATH Study website ([pathstudyinfo.nih.gov](http://pathstudyinfo.nih.gov)) 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 household anniversary month, field interviewers telephone the last known address of continuing adult SPs and parents of continuing youth SPs. This call helps to: (1) reestablish direct contact, (2) collect updated contact information, (3) identify the parent who will participate for each youth SP or shadow youth, (4) answer questions about the study, and (5) make an appointment for the in-person visit. As needed (e.g., several telephone contact attempts are unsuccessful), the field interviewers make the first direct contact in person.

Refusal letters and follow-up contacts are used with adults who are reluctant to participate and/or are located in limited-access situations (e.g., doorman buildings and gated communities). The letters are tailored to the reason for reluctance or refusal (e.g., too busy or concerned about confidentiality) and to the type of respondent (e.g., adult SP or parent of youth SP). (See Attach18.Field Data Collection Materials for an example of a refusal letter.). These letters may be sent via FedEx or priority mail to reinforce the importance of participation. Also, email messages based on refusal letters may be sent to refusing or difficult-to-contact adults who agree to be contacted this way. After the letters are sent, field interviewers work with their supervisors to tailor the approach for making follow-up contacts with the study participants by telephone or in person. Following best practice, the field interviewer making the follow-up contact is often different from the field interviewer who made the original contact.

Additional tools and approaches will be used by the PATH Study to help maximize the biospecimen response rates for Wave 4, 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 all youth SPs, new adults, and aged-up adults, the PATH Study presents information to respondents in phases to help minimize the amount of information to be simultaneously considered before consenting.<sup>12</sup> This approach includes providing information about the interview immediately prior to requesting consent for the interview, and providing information on biospecimen collection shortly before requesting consent for biospecimen collection. Moreover, because biospecimen collection follows completion of the interview, this approach allows additional time for the development of rapport, trust, and comfort between the interviewer and the respondent, which positively influences 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 professionally-formatted consent pamphlets with messages that emphasize the importance of the respondent’s contributions of biospecimens to the scientific success of the study.
- **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

---

<sup>12</sup> For the parents of continuing youth, the field interviewer requests permission for youth urine collection after reviewing the elements of permission obtained at a previous wave. For the parents of new youth and aged-up youth, the field interviewer requests permission for youth participation in the PATH Study and for youth urine collection at the same time.

and comfort with the study protocol, including obtaining consent, averting refusals, and collecting biospecimens.

- **Streamline Biospecimen Collection Procedures.** The PATH Study follows the same procedure used in previous waves of having the field interviewer collect a urine sample at the time of the interview for participants providing urine at Wave 4. Rapport that develops between the interviewer and respondent appears to have a positive influence on the respondent's willingness to provide the biospecimens. As noted, these participants will include youth, new adults, aged-up adults, and a subsample of continuing adults who initially provided urine at a previous wave and are asked to provide urine at Wave 4.
- **Enhance Quality Control.** The PATH Study data collection quality control procedures include closely monitoring interviewer-by-interviewer consent and collection rates for biospecimens, using computer-assisted recorded interviewing 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 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, cases are released to field interviewers in sets every few months; the timing of the releases is tied to the household anniversary of the SPs. 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 releases in this manner allows field interviewers the opportunity to implement the full contact strategy, including nonresponse conversion as needed, while respecting the household anniversary of the SPs.

Nonresponse 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 adjusted data set for addressing the major objectives of the study.

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

The PATH Study Wave 1 household screening and the Wave 3 data and biospecimen collection, which is currently underway, serves as an informal test of many of the methods and materials planned for Wave 4. This is reasonable given that many of the methods and materials from prior waves will be similar for Wave 4. For example, for the aged-up participants (youth SPs who age up to the adult cohort and shadow youth who age up to the youth cohort), several of the Wave 4 consent, parental permission, and youth assent procedures and methods are like those used in Wave 3. The new consent materials on collecting urine from youth have undergone cognitive testing and been revised based on the testing results. The Wave 1 methods for administering the household CAPI household screener are similar to the methods planned for Wave 4; and the Wave 4 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 Wave 3.

In addition to the informal test described in the previous paragraph, the PATH Study developed the aforementioned Interim Report based on the complete Wave 2 data and biospecimen collection and on approximately the first 6 months of Wave 3 data and biospecimen collection. Separately for these two waves, the report includes the actual or projected response rates (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 the study. Response rates are compared

throughout this report to corresponding rates projected for the entire sample, provided in Supporting Statement B of the PATH Study revision requests for the Wave 2 (OMB number 0925-0664, expiration date 9/30/2016) and Wave 3 (OMB number 0925-0664, expiration date 8/31/2018) data and biospecimen collections.

The PATH Study’s Wave 3 Interim Report is provided in Attach20.PATH Study Interim Report. Select findings on Wave 2 and Wave 3 are summarized in the remainder of this section.

### **B.4a Wave 2 Data and Biospecimen Collection**

All but one of the Wave 2 responses rates for interviews are lower than the projected response rates provided in the revision request to OMB for Wave 2; however, the response rates for biospecimen collections exceed the projected response rates. (See Table B-8.) The response rate for Wave 2 continuing adults was about three percentage points lower than projected in the revision request. However, the response rate for continuing youth was almost the same as the projection (lower by about 0.5 percentage points). The response rate for aged-up adults was about one percentage point higher than the projected rate, and the response rate for aged-up youth was approximately six percentage points lower than the projected rate. The response rates for the biospecimen collections in Wave 2 were all higher than projected. The largest differential response rates were for the tobacco use status of aged-up adults asked to provide urine and blood specimens (for each biospecimen, about eight percentage points higher for current established tobacco users than for those who were not).

Table B-8. Summary of PATH Study Wave 2 response rates

<b>Group</b>	<b>Unweighted response rate</b>	<b>Weighted response rate</b>	<b>Projected response rate*</b>
Continuing adults, Adult Interview	82.6%	83.1%	86%
Continuing youth, Youth Interview	88.5%	88.4%	90%
Aged-up adults, Adult	85.9%	85.7%	85%

<b>Group</b>	<b>Unweighted response rate</b>	<b>Weighted response rate</b>	<b>Projected response rate*</b>
Interview			
Aged-up youth, Youth Interview	82.0%	82.1%	88%
Continuing adults, urine collection	96.4%**	-	80%
Aged-up adults, urine collection	82.8%**	-	69%
Aged-up adults, blood collection	47.4%**	-	45%

\* Provided in the revision request to OMB for Wave 2 data and biospecimen collections.

\*\* See the Wave 3 Interim Report for information on why only unweighted response rates were computed.

However, nonresponse bias analysis found that many characteristics of Wave 2 respondents aligned with those of Wave 2 nonrespondents. Some exceptions were found when comparing estimates for continuing adults (current established tobacco use was lower overall, and for males, 18 to 44 year-olds, and non-Hispanic Whites among respondents) and for continuing youth (ever use of tobacco was lower overall, and for females, 14 to 17 year-olds, and non-Hispanic Whites among respondents). For continuing adults, males, 18 to 24 year-olds, and those with high school education were underrepresented among respondents; and 45 to 64 year-olds, persons with health insurance, and those with at least a bachelor's degree were overrepresented among respondents. However, estimates of cigarette smoking among adults in Wave 2 were within the range of estimates found by other national health studies.

Moreover, when the estimates of Wave 1 characteristics based on the full Wave 2 sample were adjusted for nonresponse using the Wave 2 final weights, they were almost identical to the estimates based on the Wave 1 sample and Wave 1 final weights. The Wave 2 adult cigarette smoking rates remained essentially the same using the Wave 2 final weights (compared to using the Wave 1 IPS weights), but the ever-tried cigarette smoking rates for youth remained lower than those found by other national studies. Among aged-up adults, current established tobacco users were more likely to provide urine and blood specimens; urine collection rates were very high among continuing adults.

## B.4b Wave 3 Data and Biospecimen Collection

The PATH Study Wave 3 data collection is ongoing, so response rates were calculated based on data collected from a probability subsample of cases between October 19, 2015 and April 29, 2016, using predicted response propensities for nonfinalized or interim cases. With one exception, the predicted Wave 3 responses rates exceed the projected response rates provided in the revision request to OMB for Wave 3. (See Table B-9.) The predicted Wave 3 response rate for continuing adults who responded at Wave 2 is about five percentage points higher than projected in the revision request for Wave 3; the predicted response rate for continuing youth who responded at Wave 2 is about 1.5 percentage points higher than the projection. The estimated response rate for aged-up adults who completed a youth interview at Wave 2 is seven percentage points higher than the projected rate, and the estimated response rate for aged-up youth who participated as shadow youth at Wave 2 is one percentage point lower than projected. The Wave 3 response rates for biospecimens also approximately equal or exceed the projected rates.

Table B-9. Summary of PATH Study predicted response rates for Wave 3

<b>Group</b>	<b>Unweighted predicted response rate</b>	<b>Weighted predicted response rate</b>	<b>Projected response rate*</b>
Continuing adults, Adult Interview	90.9%	91.8%	86%
Continuing youth, Youth Interview	92.5%	92.5%	91%
Aged-up adults, Adult Interview	94.4%	94.1%	87%
Aged-up youth, Youth Interview	88.0%	88.1%	89%
Continuing adults, urine collection	96.8%**	-	97%
Aged-up adults, urine collection	87.1%**	-	83%
Aged-up adults, blood collection	47.9%**	-	43%

\* Provided in the Revision Request to OMB for Wave 3 data and biospecimen collections.

\*\* See the Wave 3 Interim Report for information on why only unweighted predicted response rates were computed.

No evidence of nonresponse bias was found at Wave 3 for the Wave 1 youth or shadow youth. For Wave 1 adults, the estimated percentage of persons with at least a college degree is higher when calculated from the respondents than from the finalized or provisional nonrespondents.<sup>13</sup> The estimated percentages of males and persons age 65 or older tend to be lower for the Wave 3 respondents than for both nonrespondent groups. Current established use of tobacco is significantly lower among respondents than among provisional nonrespondents, particularly for non-Hispanic Whites and adults ages 18 to 24 at Wave 1 (for whom tobacco use rates are also significantly different when respondents are compared to finalized nonrespondents). However, as noted, these findings are preliminary pending finalization of interim cases and the remainder of data collection in Wave 3.

## **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 Attach21.List of Statistical Consultants.

---

<sup>13</sup> Provisional nonrespondents are defined to be the set of finalized nonrespondents plus interim refusals and persons who are difficult to locate.

## References

- Brick, J.M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*, 29, 329–374.
- Contoyannis, P., Jones, A.M., & Rice, N. (2004). The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics*, 19, 473–503.
- Guilliford, M.C., Ukoumunne, O.C., & 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.
- Kalton, G. (1986). Handling wave nonresponse in panel surveys. *Journal of Official Statistics*, 2, 303–314.
- Kashihara, D., & 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.
- Lundström, S., & Särndal, C-E (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, 15 (2), 305–327.
- Lynn, P. (2006). Quality Profile: British Household Panel Study. Essex, UK: Institute for Social and Economic Research. Retrieved from <https://www.iser.essex.ac.uk/files/bhps/quality-profiles/BHPS-QP-01-03-06-v2.pdf>
- 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. Retrieved 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. Retrieved 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.

- Roth, S., Montaquila, J., & Chapman, C. (2006). Nonresponse bias in the 2005 National Household Education Surveys Program. Technical Report. (NCES 2007-016). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Schouten, B., & de Nooij, G. (2005). *Nonresponse adjustment using classification trees*. Discussion paper 05001, Statistics Netherlands. Available at [www.cbs.nl](http://www.cbs.nl).
- Särndal, C.E., & Swensson, B. (1987). A general view of estimation for two phases of selection with applications to two phase sampling and non-response. *International Statistical Review*, 55, 279-294.
- Thompson, D.M., Fernald, D.H., & Mold, J.W. (2012). Intracluster correlation coefficients typical of cluster-randomized studies. *Annals of Family Medicine*, 10, 235-240.