NSDUH MENTAL HEALTH LONGITUDINAL FEASIBILITY STUDY

Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Rockville, Maryland 20857

September 2011

This page intentionally left blank

NSDUH MENTAL HEALTH LONGITUDINAL FEASIBILITY STUDY

Contract No. 283-2004-00022
RTI Project No. 0212682.120.008.002.004

RTI Authors:

Cristie Glasheen
Rebecca Granger
Rhonda Karg
Patricia LeBaron
Katie Morton
Heather Ringeisen
Leyla Stambaugh
Kevin Wang
David Wilson

RTI Project Director:

Thomas G. Virag

SAMHSA Authors:

Peggy Barker
Jonaki Bose
Joe Gfroerer
Art Hughes
Dicy Painter

SAMHSA Project Officer:

Peter Tice

For questions about this report, please email Peter.Tice@samhsa.hhs.gov.

Prepared for Substance Abuse and Mental Health Services Administration, Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

September 2011

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2011). *NSDUH Mental Health Longitudinal Feasibility Study*. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Acknowledgments

This summary was prepared for the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), formerly the Office of Applied Studies (OAS), under Contract No. 283-2004-00022, by staff at RTI International (a trade name of Research Triangle Institute). In addition to the listed authors, significant contributors include Jeremy Aldworth, Patrick Chen, Jim Chromy, David Hunter, Doug Currivan, Phillip Kott, Larry Kroutil, Mary Ellen Marsden, Marty Meyer, Michael Pemberton, Scott Novak, and Tom Virag.

Table of Contents

Chapter **Page**

1. Introduction 1

2. Background 3
   - 2.1 Epidemiologic Catchment Area Study 3
   - 2.2 National Comorbidity Study 4
   - 2.3 National Epidemiological Survey of Alcohol and Related Conditions 4

3. Issues for Consideration 7
   - 3.1 Measures of Interest 7
     - 3.1.1 Prevalence 7
     - 3.1.2 Incidence 7
     - 3.1.3 Course 8
     - 3.1.4 Comorbidity 8
     - 3.1.5 Risk and Protective Factors 8
   - 3.2 Design Components 9
     - 3.2.1 Study Population 9
     - 3.2.2 Design Structure 10
     - 3.2.3 Periodicity 13
   - 3.3 Questionnaire Issues 15
     - 3.3.1 Domains of Interest 15
     - 3.3.2 Dimensional and Categorical Approaches to Assessing Mental Disorders 16
     - 3.3.3 Scope of the Assessment 16
     - 3.3.4 Psychometric Properties 17
     - 3.3.5 Instrument Format 17
     - 3.3.6 Instrument Comparability 18
     - 3.3.7 Candidate Instruments 19
     - 3.3.8 Statistics of Test Performance 32
   - 3.4 Sample Issues 33
     - 3.4.1 Sampling Strategy 34
     - 3.4.2 Sample Design Advantages and Disadvantages 34
     - 3.4.3 Sample Size 35
   - 3.5 Data Collection Implications 45
     - 3.5.1 Interviewer Type and Mode of Administration 45
     - 3.5.2 Recruiting and Training Interviewers 46
     - 3.5.3 IRB and OMB Issues 48
     - 3.5.4 Tracing and Panel Maintenance 49
     - 3.5.5 Respondent Incentives 51
   - 3.6 Data Quality 53
     - 3.6.1 Reference Periods 53
     - 3.6.2 Bounding 54
     - 3.6.3 Panel Conditioning 54
     - 3.6.4 Seam Effects 56
     - 3.6.5 Coverage and Response Rate Issues 56
   - 3.7 Trend Analysis 57

3.8 Effect on Main Study 57
3.9 Dissemination 59

4. Longitudinal Design Scenarios 61
4.1 Design Scenarios 61
4.2 Instrument Options 62
4.3 Sample Sizes 64
4.4 Cost Estimates for Design Scenarios 66

5. Conclusion 71

6. References 73

Appendix

Appendix A: Longitudinal Design Scenario Diagrams A-1

List of Tables

Table **Page**

1. Sequential Cohort Design Example for Persons Aged 18 to 22 at Baseline 11
2. Adult Mental Health Instruments.....20
3. Mental Health Surveillance Study Measures by Domain 24
4. Instruments for Assessing Child and Adolescent Mental Health 27
5. Adult Disorders by Prevalence Group 36
6. Youth-Specific Disorders by Prevalence Group.....36
7. Design Effects from the 2010 Mental Health Surveillance Study SCID Data (18 or older) 38
8. Adult Cohort: Low Prevalence Disorders (Adjustment Disorders, Bipolar, Psychotic Disorders), $p = .03$ 40
9. Adult Cohort: Medium Prevalence Disorders (Eating Disorders, Major Depressive Disorder), $p = .07$ 41
10. Adult Cohort: High Prevalence Disorders (Anxiety Disorders, Impulse-Control Disorders, General Mood Disorders), $p = .25$.....42
11. Youth Cohort: Low Prevalence Disorders (Attention-Deficit and Disruptive Behavior Disorders, Mood Disorders), $p = .03$.....43
12. Youth Cohort: High Prevalence Disorders (Any Mental Health Disorder), $p = .2$ 44

13 Add Health Wave III and Wave IV Comparison—Percentage Located, by Tracing Step 50

14. Tracing Strategies for Minimizing Attrition 51
15. Advantages and Disadvantages of Instrument Options 63
16. Adult Cohort: Assumed Response Rates and Sample Sizes for Longitudinal Design Scenarios A, B, and C 65
17. Youth Cohort: Assumed Response Rates and Sample Sizes for Longitudinal Design Scenarios A, B, and C.....65
18. Cost Estimates for Longitudinal Design Scenarios 68
19. Cost Estimate by Year for Scenarios A, B, and C for the SCID/K-SADS Instrument Option 69

List of Figures

Figure **Page**

1. Longitudinal Design Scenarios 61
2. Marginal Cost Curve Using the SCID/K-SADS Option for Scenario A 67

1. Introduction

The National Institute of Mental Health (NIMH) has asked the Substance Abuse and Mental Health Services Administration (SAMHSA) to explore whether a nationally representative longitudinal mental health component for adults and youths can be included as part of the National Survey on Drug Use and Health (NSDUH). The main analytic goal of a mental health longitudinal component to NSDUH would be to understand the trajectories of mental health problems. These include onset, persistence, remission, and relapse, as well as comorbidity and predictors related to mental health problems. Various longitudinal study designs would permit estimation of the prevalence, incidence (new cases), or course (e.g., duration, number of episodes) of mental disorders. While there are national published estimates of the prevalence of mental disorders for youth and adults, only sparse information is available about the incidence of mental health conditions. Even less is known about the long-term course of mental disorders, including trajectories of persistence, remission, and relapse. There are very few longitudinal studies focused on the course of mental disorders. Most of these are small and composed of persons in treatment. Estimation of the incidence of mental disorders would also be particularly relevant for youths because a number of mental disorders typically start during early adolescence or young adulthood.

The analytic goals and measures of interest for a mental health longitudinal component, as well as the impact on the NSDUH main study and other design and cost issues, must be considered prior to the implementation of such a study. This report describes these issues along with three design scenarios developed to meet the analytic goals of a NSDUH mental health longitudinal study. Section 2 includes a summary of relevant mental health longitudinal studies previously conducted in the United States. Section 3 describes the issues that need to be considered prior to the implementation of any mental health longitudinal component to the NSDUH. Section 4 provides three longitudinal design scenarios and cost estimates for conducting each design. Section 5 provides conclusions of the report.

2. Background

Of the mental health longitudinal studies that were evaluated for relevance to the NSDUH, a few were identified as particularly noteworthy with regard to mental health considerations. This section provides background on the Epidemiologic Catchment Area (ECA) study, the National Comorbidity Study-2 (NCS-2), and the National Epidemiological Survey of Alcohol and Related Conditions (NESARC). Research from these studies is referenced throughout the report.

2.1 Epidemiologic Catchment Area Study

One of the earliest large-scale population-based studies of mental illness in the Unites States is the ECA study, sponsored by the National Institute of Mental Health (NIMH). Starting in 1980, the ECA project collected data on the prevalence and incidence of mental illness from a general sample of the adult U.S. population (Regier et al., 1984). Recruitment was completed through a multistage probability sample of adults living in households in five catchment areas (New Haven, CT; Baltimore, MD; St. Louis, MO; Durham, NC; and Los Angeles, CA) and followed over 20,000 participants, longitudinally. The core ECA project consisted of a baseline and a 12-month follow-up interview conducted in the home; a 6-month telephone follow-up was also completed. Some sites have had additional follow-ups, but these were not standardized across cities and, therefore, were not relevant enough for the current consideration of representative mental health studies. This section focuses only on the core study.

The ECA interview used the fully-structured Diagnostic Interview Schedule (DIS) version II (in New Haven) and III (in all other cites) and assessed Diagnostic and Statistical Manual of Mental Disorders (DSM) III diagnoses of single-episode major depression, recurrent major depression, dysthymia, manic episodes, and bipolar and atypical bipolar disorder; schizophrenia and schizophreniform disorder; alcohol and drug abuse or dependence; phobia, panic, and obsessive compulsive disorder; somatization disorder; antisocial personality disorder; and anorexia nervosa (Robins, Helzer, Croughan, & Ratcliff, 1981). The use of a standardized diagnostic system assisted in providing prevalence estimates that could be compared over time and across sites. Inclusion of two additional follow-up periods strengthened the 1-year prevalence estimates by catching cases at multiple time points throughout the year. With a single assessment, respondents who had subthreshold symptoms at baseline would not have been included in 12-month prevalence estimates, even if their symptoms exceeded threshold levels later in the year. Furthermore, the assessment of incident cases (new cases) was possible only because a minimum of three time points (at baseline and at 6 and 12 months) were collected. This enabled incidence rates of various disorders to be calculated within specific groups, like age groups and race. Despite these strengths, the ECA study is now over 30 years old, diagnostic criteria have been revised, and psychiatric assessment instruments have improved (Robins & Cottler, 2004), Furthermore, with only 1 year of follow-up, trajectories of mental illness in the population could not be fully enumerated.

2.2 National Comorbidity Study

The National Comorbidity Study (NCS)[1] was the first nationally representative study to use a fully structured interview to examine the prevalence and correlates of mental illness in the United States (Kessler & Wang, 2008). The NCS was conducted from 1990 to 1992 and surveyed a noninstitutionalized population aged 15 to 54 in the 48 contiguous states using a stratified, multistage area probability sample. Beginning in 2001, the NCS-2 follow-up face-to-face survey was conducted with the same respondents using computer-assisted personal interviewing (CAPI) technology, which took between 90 minutes and 5 hours to complete. The purpose of the 10-year follow-up was to measure changes in mental health status and substance use disorders and the link between primary mental disorders and subsequent substance use disorders. Over 8,000 respondents completed the NCS, with a response rate of 82.6 percent. The NCS-2 successfully traced 5,463 respondents, of whom 166 were deceased and 5,001 re-interviewed, for an unweighted response rate conditional on participation in the original NCS of 87.6 percent and an unweighted unconditional response rate of 72.2 percent.

The NCS used the World Health Organization's Composite International Diagnostic Interview (CIDI), which is a fully structured diagnostic interview based upon the previously developed Diagnostic Interview Schedule (Robins et al., 1988). The CIDI's fully structured nature enables nonclinicians to administer the interview and still derive reliable diagnoses. The NCS and the NCS-2 instruments collected data about current and lifetime (retrospective) prevalence of mental illnesses and service use in various treatment sectors (Kessler & Merikangas, 2004). Data were collected on some of the most common diagnoses, as well as the most burdensome disorders on individuals and society. Disorders assessed included major depression, mania, dysthymia, panic disorder, social phobia, agoraphobia, simple phobia, generalized anxiety disorder, alcohol abuse and dependency, drug abuse and dependency, antisocial personality disorder, and nonaffective psychosis.

The NCS improved on the ECA in a number of ways (e.g., using a nationally representative sample, assessing children as young as 15, using an improved clinical assessment measure, and gathering data on comorbidities and correlates); however, the length of time between assessments (10 years) and the number of follow-ups limit the depth of information that could be attained. Although an overall level of change in diagnostic prevalence could be assessed, the patterns of those changes could not be observed with only a baseline and a single follow-up period. For example, individuals obtaining a diagnosis at baseline might have a recurrent course with periods of remission and recurrence throughout the 10 years between the baseline and follow-up. These periodic changes could be captured only retrospectively at the follow-up period, which is subject to a number of recall-related biases, including differential recall in cases and controls and faulty recall in general or for older events (Coughlin, 1990).

2.3 National Epidemiological Survey of Alcohol and Related Conditions

The National Epidemiological Survey of Alcohol and Related Conditions (NESARC) is a longitudinal study of over 49,000 respondents measuring a number of mental illnesses, including past year affective disorders, anxiety disorders, substance use disorders, and lifetime personality

[1] http://www.hcp.med.harvard.edu/ncs/

disorders (Grant, Kaplan, Shepard, & Moore, 2003; Grant & Kaplan, 2005). The NESARC improved on the ECA with a nationally representative survey design and improved on the NCS-2 with a shorter follow-up period. It also assessed a wider variety of disorders than both the ECA and the NCS. First conducted in 2001, the NESARC surveyed noninstitutionalized, adult civilians in the United States, including the District of Columbia, Alaska, and Hawaii. Much like the NSDUH, the NESARC sampled people living in a variety of group accommodations, such as boarding or rooming houses and college quarters. Military personnel living off base were also eligible for the study. The sampling frame for the NESARC was obtained from the housing units (HUs) for the 2000 to 2001 United States Census. To obtain a clearer picture of mental illnesses in young adults, blacks, and Hispanics, these populations were oversampled. The initial unweighted household response rate was 89 percent, while the unweighted person response rate was 93 percent. Respondents were reinterviewed 3 years later with an unweighted response rate of 86.7 percent.

The NESARC interview obtained diagnoses of major mental disorders using the Alcohol Use Disorder and Associated Disabilities Interview Schedule, version IV (AUDADIS-IV), which is a fully structured, diagnostic CAPI lasting about 1 hour that captures major DSM-IV axis I (clinical) disorders and axis II (personality) disorders (Grant, Kaplan, Shepard, & Moore, 2003). Specifically, the NESARC measures past year and lifetime alcohol, tobacco, and drug use disorders, major depressive disorder, dysthymia, manic episodes, hypomanic episodes, panic disorder with and without agoraphobia, agoraphobia without history of panic disorder, social phobia, specific phobia, generalized anxiety disorder, and pathological gambling. Lifetime estimates of avoidant personality disorder, dependent personality disorder, obsessive-compulsive personality disorder, paranoid personality disorder, schizoid personality disorder, histrionic personality disorder, and antisocial personality disorder were collected. Wave 2 borderline, narcissistic, and schizotypal personality disorder were also collected. The Wave 2 instrument also measured post-traumatic stress disorder and attention deficit hyperactivity disorder (Grant, Dawson et al., 2003).

The NESARC built upon the NCS in a number of important ways, including sampling those in group housing units, oversampling minorities, having a shorter follow-up period to decrease recall bias, and obtaining national estimates for number of personality disorders never previously assessed. However, two points were not addressed. First, the NESARC consisted of only one follow-up period and thus did not address the question of symptom patterns over time. Second, unlike NCS, the NESARC did not include anyone under the age of 18 despite the high prevalence of mental illness in children (Ringel & Sturm, 2001).

3. Issues for Consideration

This section provides a discussion of the issues to be considered in an evaluation of the feasibility of including a mental health longitudinal component in the NSDUH.

3.1 Measures of Interest

A number of main questions are of interest when studying mental disorders, including the following:

- How many people in the population have or have ever had a mental illness (past year, current, and lifetime prevalence)?
- How many new cases of mental illness arise annually in the population (annual incidence)?
- When individuals have multiple disorders, what types tend to co-occur and which usually present first (incidence of comorbidity)?
- When individuals have a mental disorder, what is the course (chronicity, remission, and relapse) of that disorder and does it differ by characteristics such as treatment or comorbidity?
- What descriptive and predictive characteristics are associated with the incidence and course of mental illness?

Each of these questions requires individual consideration when developing a study design.

3.1.1 Prevalence

Prevalence is the most direct measure of mental illness available. It is the number of *existing* cases adjusted for the size of the population under study. As such, no longitudinal component is needed to collect prevalence data; rather, a single cross-sectional assessment is sufficient. Currently, a few studies have produced national prevalence estimates for some mental illnesses, including NSDUH, the National Comorbidity Study Replication (NCS-R), and the National Epidemiologic Study of Alcohol and Related Conditions (NESARC).

3.1.2 Incidence

The incidence rate is a ratio of the number of *new* cases in a population to the size of the population of interest over a specified period of time. It is possible to derive mental illness incidence estimates from a single cross-sectional assessment by retrospectively assessing the age of onset for a new disorder; however this method is subject to a great deal of recall bias for longer recall periods (Johnson & Schultz, 2005; Prusoff, Merikangas, & Weissman,1988). Longitudinal assessments with a baseline and a single follow-up are more appropriate for obtaining incidence estimates. The baseline measure allows for identifying individuals who already have a mental illness (prevalent cases), enabling them to be differentiated from the new

cases (incident cases) that develop during the period of observation. Very few national studies exist that provide information on incidence of mental illness. The NESARC gathered incidence data over a 3-year period for several disorders, although the primary goal of the study was to examine alcohol use disorders.

3.1.3 Course

Measuring the course of a mental illness can be done at a group level (e.g., among people with depression, what is the average duration of clinical-level symptoms) or within an individual (e.g., what is the pattern of symptoms within an individual). Identifying the course of a disorder requires a longitudinal study design because it is a measure of change over time.

Study designs that measure the course of disorders must consider the natural pattern of disorders in the population. (Unfortunately, the data on natural course are sparse for most mental disorders.) For example, some disorders may be short lived in the majority of people (e.g., depression [Kessler, Zhao, Blazer, & Swartz, 1997]), and therefore, assessments that occur with too long of an interval in between may miss the timing of remission. On the other hand, disorders that are chronic in nature will show little variability over short-term assessments, and if follow-up is discontinued too early, remission or relapse may be missed. For many disorders, there may be a point of diminishing returns where longer follow-up yields less and less information because cases most likely to remit or relapse would have already done so. It will be necessary to consider these competing needs in any longitudinal study assessing multiple disorders.

Studies of pattern are best done with at least three assessment points, and recurrence can be identified only if at least three assessments are conducted. With only two assessments, variation in disease pattern may be missed because the only pattern identifiable with two data points is a straight line (consistently improving, consistently worsening, or remaining stable). Moreover, recurrence implies a period of remission. Therefore, any study that has only two time points can identify only three outcomes: the development of a disorder, the maintenance of a disorder, or the remission of a disorder. To identify recurrence, a third time-point must be added so there is the opportunity measure improvement and then relapse.

3.1.4 Comorbidity

Comorbidity (multiple disorders) among individuals with a mental illness is a common phenomenon (Kessler et al., 2005). While cross-sectional studies can assess the prevalence of co-morbidity, they are not the optimal method for determining the sequence of disorder onset. Prospective longitudinal studies are better designed to determine which disorder presented first, whether there are differences in specific disorder onset patterns, and whether outcomes vary by comorbidity sequence (for example, individuals with an onset of depression first followed by anxiety may have a different outcome than individuals with anxiety presenting first and depression second).

3.1.5 Risk and Protective Factors

Epidemiological studies can be primarily descriptive or can also collect predictive information. Descriptive studies collect information about risk indicators, also called risk markers, which are characteristics that are associated with mental illness risk but do not

necessarily cause mental illness (e.g. race, sex, and age). Predictive studies, by contrast, collect information on causal risk factors as well as risk indicators. Risk factors for mental health problems are often best examined through longitudinal studies. Assessing a causal relationship between an exposure and an outcome requires being able to verify that the exposure occurred prior to the outcome, something that often cannot be determined in cross-sectional studies where the exposure and outcome are measured simultaneously. However, some considerations must be made for assessing risk factors in longitudinal studies, including the presence of recall bias. For example, while young adults may be better able to recall events that happened during childhood, asking the same of older adults may be prone to more recall errors, although significant life events (such as experiencing abuse or the parents' divorce) will be subject to less recall bias overall. One advantage to longitudinal designs is the ability to look at proximal risk factors and their effect on outcomes in subsequent waves—for example, high alcohol use at baseline and the development of a mental illness by the completion of the study.

Assessment of protective factors, such as the effect of obtaining treatment on mental illness course is similar to that of measuring risk factors. Shorter recall periods are less subject to subjective memory biases. For instance, people who obtained treatment 30 years prior may recall treatment being more or less successful than they would have reported at the time, depending on whether they experienced a relapse in symptoms later. These are challenges existent in all longitudinal designs, and a balance must be obtained between too frequent assessments that might reduce participation and yield only small amounts of data because there was little variation over the short-term follow-up, and too few assessments at too great an interval that might interfere with recall and reporting.

3.2 Design Components

This section describes key design components for consideration, including the study population, design structure, and periodicity of assessments.

3.2.1 Study Population

Most symptoms of mental disorders begin prior to adulthood (Kessler et al., 2005), so the inclusion of a youth component is vital to examining trajectories of mental illness in the United States. When considering a mental health longitudinal component for NSDUH, youths aged 12 or older could be sampled at the baseline assessment for longitudinal follow-up. NSDUH currently does not include youths younger than 12; however, epidemiological data suggest that by the age of 12, mental disorders of childhood (e.g., adjustment disorders and enuresis) have started to wane and disorders of adolescence (e.g., conduct disorder, mood disorders, substance use disorders) have not yet emerged (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003). From this perspective, the age of 12 may be an acceptable starting point for a longitudinal study of incidence and course of mental disorders during adolescence.

Longitudinal studies must also weigh the cost and benefits of following up groups of people in whom new cases of mental disorders are rare. This is particularly relevant with regard to the upper age limits for inclusion in the study. Most mental disorders that onset in adulthood, have onsets in early adulthood. For example, some eating disorders (Hudson, Hiripi, Pope, & Kessler, 2007) and bipolar disorders (Merikangas et al., 2011) have mean ages of onset in the

early to mid-20s, with the distribution of onset tapering off by the mid-30s. This suggests that the majority of information gathered about these disorders will be obtained from the younger age groups.

However, the decision to exclude older individuals must be considered carefully, because although the majority of disorders are going to be obtained early in adulthood, some disorders have a wider distribution of onset ages. This is particularly true for major depression, generalized anxiety disorder, and schizophrenia, all of which have a substantial number of onsets in later ages. For example, one population-based study found a 1 percent increase in schizophrenia prevalence between the ages 30–44 and 45–54, suggesting a considerable number of new cases were arising in the population during those years (Perala et al., 2007). Likewise, the prevalence of generalized anxiety disorder increased from 6.8 percent for ages 30–44 to 7.7 percent for ages 45–59 in the NCS-R (Kessler et al., 2005). The Baltimore follow-up of the Epidemiologic Catchment Area study reported an annual incidence of major depression of 5.4 per 1,000 between the ages of 30 and 44, suggesting that not including these age groups would miss a substantial number of cases and limit generalizability in older ages (Eaton, Kalaydjian, Scharfstein, Mezuk, & Ding, 2007). Similar findings for mid-life major depression onset were found in the NESARC study (NIH/National Institute on Alcohol Abuse and Alcoholism, 2005).

3.2.2 Design Structure

First described by Bell (1953), the sequential cohort study design was developed to combine the benefits of cross-sectional and longitudinal studies while overcoming the weaknesses of both designs. A sequential cohort study design is also referred to in the literature as "sequential cohort" (Happ, Sereika, Garrett, & Tate, 2008), "accelerated longitudinal" (Bell, 1953; Miyazaki & Raudenbush, 2000), "cross-sequential" (Miyazaki & Raudenbush, 2000), and "overlapping cohort" (Raudenbush & Chan, 1993). This design has been used in several notable studies, including the National Youth Survey (Elliott, Huizinga, & Menard, 1989), the Project on Human Development in Chicago Neighborhoods (Sampson, Raudenbush, & Earls, 1997), and the Great Smoky Mountains Study, a regional longitudinal study of children's mental health (Costello et al., 1996).

The sequential cohort design is similar to a standard longitudinal study, varying only in the timing of the follow-up waves and the use of multiple age groups within a sample. A traditional cohort design focusing on examining the development of mental illnesses across the lifespan, for example, might sample a representative group of 18 year olds and follow them periodically until they reach age 54 to estimate the incidence of mental disorders across these ages (e.g., The Finland Birth Cohorts [Tiihonen, Isohanni, Rasanen, Koiranen, & Moring, 1997]). While informative, this design has two major limitations. First, the length of follow-up is considerable. This adds substantial management and participant burden and leads to significant loss to follow-up in later years. Second, the estimates from a cohort of a single age group may be subject to period and cohort effects. These biases arise when the participants in a cohort study all share a single exposure (such as year of birth or exposure to major events, like a natural disaster). Since all of the members of the cohort experience these events, it is not possible to know if the estimates from that cohort would generalize to others not born in that year or not exposed to that event.

A sequential cohort study is designed to address these issues by reducing follow-up length, while keeping the breadth of the study the same and identifying and controlling for period and cohort effects. In essence, an accelerated cohort is composed of different age groups, followed in the same manner as a traditional longitudinal study, but with overlap points that allow estimates to be combined across all the ages contained in the cohort. For example, instead of following a group of 18 year olds until they reach age 54, the assessment covers participants aged 18 to 52, with a baseline and two annual follow-ups, to produce data covering ages 18–54 in the span of 2 years.

The process of conducting a sequential cohort study is essentially the same as conducting a traditional longitudinal study. A group of participants is recruited and followed for a specified period of time. The differences are (a) the deliberate inclusion of different age groups at baseline and (b) the minimum number of follow-ups to allow for linking the different age groups.

Since NSDUH already conducts sampling across ages to ensure a population-representative sample, the different ages at baseline are already incorporated into the sampling design, making the sequential cohort design a natural fit for these research purposes. The only additional requirement is additional follow-up assessments (which, for statistical purposes is three [Miyazaki & Raudenbush, 2000]).

To illustrate, consider participants aged 18–22 at the beginning of the study with three primary assessments: a baseline and two follow-ups. Although the assessments all occur at the same calendar-time for all participants (baseline and years 1 and 2 of the study), the age of participants at each assessment naturally varies. Some participants may be 18 at baseline whereas others may be 22. Table 1 shows participant age at each follow-up for participants who were aged 18–22 at baseline. This table demonstrates that, while only two calendar years of time pass between the baseline and the second annual follow-up, the data collected cover 6 years of age (18–24). Linking these age groups is possible because of the two overlap periods within each age cohort. For example, the participants who are aged 18 at baseline have overlapping assessments at age 19 and 20 with participants who were age 19 or 20 at baseline.

Table 1. Sequential Cohort Design Example for Persons Aged 18 to 22 at Baseline

Age at Baseline	Aged 18	Aged 19	Aged 20	Aged 21	Aged 22	Aged 23	Aged 24
18	X	X	X				
19		X	X	X			
20			X	X	X		
21				X	X	X	
22					X	X	X

The additional benefit is that the inclusion of different age groups in the study allows for evaluating and even controlling for period, age, and cohort effects. It should be noted that controlling for all potential period or cohort effects is not always possible in this design, but it is never possible in a traditional single age-based cohort study.

Taken as a whole, a sequential cohort design can produce estimates of the incidence, prevalence, and comorbidity of mental illness for persons aged 18 to 54 in a period of 3 years,

thus saving time, effort, and costs in comparison to following a single cohort for 36 years. The shorter time frame for follow-up should reduce sample attrition and improve participation rates. Note that due to changes in the diagnostic criteria for youths versus adults, the youths cannot be included in the estimates generated for adults and therefore have a more traditional longitudinal design. Additionally, the estimates that can be produced are dependent on both study design and sample sizes, so although the design is capable of producing individual age-based estimates, the sample sizes proposed in Section 4 are based on producing a single estimate for all adults aged 18 to 54.

Analyzing the data produced from sequential cohort designs demonstrates additional benefits to this design. The data can be analyzed using all cross-sectional and longitudinal statistical methods usually employed to answer research questions in standard cohort studies. In addition, multilevel modeling (most often Hierarchical Linear Modeling) can be used to link the age groups across the study period to test for period and cohort effects, and produce growth curves for the incidence of mental illness across the lifespan (Miyazaki & Raudenbush, 2000; Raudenbush & Chan, 1993). Thus, the sequential cohort design does not limit potential analytic methods but allows for additional tools for analysis. Moreover, because the assessments are done concurrently (all baseline assessments are completed in the same calendar period, all follow-ups the next calendar year, etc., regardless of age), the schedule of availability and presentation of the datasets (baseline plus subsequent follow-ups) are the same as a standard longitudinal study.

Because the sequential cohort design differs from a traditional longitudinal study only in the diversity of ages and number of follow-ups included, the limitations are similar to all longitudinal studies, mainly confounding due to differential participation of respondents (volunteer bias) and loss to follow-up. There is, however, one caution with a sequential cohort design. In theory, this sequential cohort study would generate data equivalent to following 18 year olds until they were 54 years of age while taking only 2 years to conduct. As mentioned previously, however, not all period/cohort effects can be controlled. Traditionally, statistical testing is done to determine whether it is reasonable to link the age cohorts together (i.e., treating the older age cohorts as if they were an older extension of the younger cohorts). Sometimes this examination determines that there are period or cohort effects influencing the estimates of certain age groups that cannot be controlled. In this case, the age groups should not be combined for a single growth curve estimate. The impact of this is minimal. All traditional longitudinal analyses within the study period can still be completed; this prevents only producing a single growth curve estimate encompassing the entire 18–54 age range, although smaller age curves might still be reasonable (for example, the discrepancy may be at later ages, so a curve can still be used to combine younger ages). This limitation also provides important information. If the outcome of interest was the incidence of mental illness, and a test of the data suggested that generating a single growth curve across the lifespan was not reasonable, this result suggests that there is a significant historical effect wherein one age group has a significantly different incidence rate than would be expected from another age group once they reached that age. If only a single age group had been followed, there would be no way to determine how much their incidence was different from others of a different age group.

3.2.3 Periodicity

This section describes four assessment types for consideration when developing a mental health longitudinal component for NSDUH: baseline assessment, annual follow-ups, 6-month follow-ups, and long-term follow-ups. The baseline assessment and annual follow-ups are required for a rigorous study design capturing information mainly on incidence and prevalence, whereas the six-month and long-term follow-ups provide additional information on course of disorder but are not necessary for incidence or prevalence.

3.2.3.1 Baseline Assessments

The baseline assessments for adolescents and adults would include a full diagnostic interview assessing lifetime and current psychiatric disorders. Obtaining lifetime occurrence of disorders at baseline is required to determine which individuals have never had a psychiatric disorder, those who have recently developed a psychiatric disorder but have never had one before (incidence), and those who are having a recurrent or continuing psychiatric disorder (prevalence and course). Getting this information is particularly important among older individuals because they are more likely to have had a prior mental illness than children.

Even among youth there may be individuals who have had a previous lifetime disorder but do not currently meet criteria for a mental disorder, and some of these will have a recurrence of the disorder at a later follow-up. Precedent for this method of assessing relapse in children is found in Lewinsohn and colleagues' (1993) longitudinal study of children's mental health in Oregon. Measuring lifetime and current disorders at baseline will also provide estimates of recent onset disorder, which provides data on incidence (number of new cases) of mental disorders for the year prior to baseline (i.e., between the ages of 11 and 12).

3.2.3.2 Annual Follow-ups

Annual follow-ups are recommended for at least 2 years after baseline to establish a sequential cohort design, enable a rich and detailed level of longitudinal data, and provide enough age overlap for statistical analysis. The annual follow-ups, regardless of age, gather information on newly developing cases of mental illness (incidence) across the lifespan while minimizing the recall bias introduced by longer intervals of follow-up. The annual assessments gather information only on current disorders and those present in the past year, as mental disorders prior to that were assessed at baseline. The Structured Clinical Interview for DSM-IV-TR (SCID) interview and comparable child diagnostic interviews such as the Kiddie-Schedule of Affective Disorders and Schizophrenia (K-SADS) or Diagnostic Interview Schedule for Children (DISC) are designed to assess past-year and current symptoms to determine diagnosis.

Annual assessments also gather data on continued course information for respondents that have already screened positive for a disorder as well as the development of comorbidity (co-occurrence of multiple disorders). Unlike studies of prevalence, the longitudinal design may help ascertain the order of the development for mental illnesses that frequently co-occur (e.g., depression and anxiety) without relying on retrospective accounts (i.e., recall for a period longer than a year), which have been shown to be biased in assessments of mental illness (Ebner-Priemer et al., 2006; Mineka & Nugent, 1995).

Annual follow-ups are also useful for gathering information on other relevant factors such as suicidal thoughts/behaviors, alcohol and substance use, mental health treatment patterns, and significant life events that may contribute to mental illness onset (e.g., divorce, job loss, school change).

Longitudinal studies of child mental health typically include annual follow-up waves. No large epidemiological studies were identified that had follow-up intervals of less than 12 months. Recent findings from the Great Smoky Mountain Study (GSMS) suggest that cumulative prevalence of mental disorders grows significantly over the course of adolescence, with more than 60 percent of the adolescents assessed having had a specific disorder assessed by the Child and Adolescent Psychiatric Assessment (CAPA) by the age of 21 (Copeland, Shanahan, Costello, & Angold, 2011). Yet, at any annual wave, prevalence ranges from 12 to 17 percent, suggesting a high incidence of new cases during adolescence. These findings suggest that annually spaced follow-up waves will enable the detection of many new cases each year, as well as tracking of continuity and discontinuity of disorders through adolescence.

3.2.3.3 Six-Month Follow-ups (Brief Screening)

The 6-month follow-ups in adolescents and adults would capture course information for various disorders, particularly those that follow an episodic pattern—for example, major depression (Kessler et al., 1997), bipolar disorders (Solomon et al., 2010), and panic disorders (Batelaan et al., 2010)—without having to rely on respondents' recall. To track the course of disorders, the initial 6-month follow-up would be required only for individuals with a psychiatric diagnosis at baseline. These assessments would also be useful for gathering information on related factors that may be transient or have a short duration, such as work or relationship stress, and on brief mental health treatment.

This process can be conceptualized as producing a subset of the overall population representative of people with a psychiatric diagnosis, with weights generated to generalize back to the national level. The standard information generated by any cross-sectional study can be produced to obtain descriptive statistics on these individuals. However, the strength of the 6-month follow-up arises when the data are combined with the annual assessments. When the 6-month assessment is combined with information from the baseline and annual follow-ups, it equals three assessments that are 6 months apart and a fourth 12-months later. These data can be combined to evaluate individual-level diagnostic trajectories, the development of comorbidities, and characteristics associated with different diagnostic trajectories. Advances in statistical methods have improved using longitudinal analyses with unequal follow-up periods at the group and individual level (Vicente & Woodworth, 1994; Thwin & Gagnon, 2008), an important advancement given the typical difficulty in completing follow-ups on everyone exactly 6 and 12 months after their baseline interviews.

Findings from studies using annual follow-up waves have revealed substantial change from year to year during adolescence, including both homotypic (continuity of the same disorder) and heterotypic (moving from one disorder to another) paths, as well as the emergence of concurrent comorbidity of two or more disorders in later adolescence (Angold, Costello, & Erkanli, 1999; Costello, Foley, & Angold, 2006; Rowe, Costello, Angold, Copeland, & Maughan, 2010), thus suggesting rapid fluctuation among children with psychiatric diagnoses.

Although no longitudinal studies with follow-up intervals less than 12 months were identified, the findings from the Great Smoky Mountain Study on incidence and course of disorder are based on current disorders (defined as disorders in the past 3 months) assessed at each annual wave. Therefore, the enormous annual changes observed in that study were found without even accounting for a full 9-month period between waves. Hence, two 6-month follow-ups are proposed for adolescents. The first includes only individuals with psychiatric diagnoses at baseline, and the second would include anyone with a psychiatric diagnosis at any of the prior assessments.

As with the 6-month follow-ups for adults, the child assessments could be used to assess changes in functional impairment or service use among youths with a mental disorder. The type of service/treatment received could be obtained, and comparisons could be made of disorder course between youths who did and did not use services. Such an assessment might also include a brief measure of functional impairment to examine the impact of mental disorder on a youth's ability to function at home, with peers, or in the community. The purpose of this 6-month assessment would not be to identify the onset of new disorders (incidence), but, rather, to track the course and impact of previously identified disorders.

3.2.3.4 Long-Term Follow-up

Long-term follow-ups would obtain further information on the course of disorders with a focus on recurrence and remission. Some psychiatric disorders (e.g., dysthymia) have chronicity as a feature of the diagnosis, and many disorders appear to follow a chronic course—for example, agoraphobia (Wittchen, Gloster, Beesdo-Baum, Fava, & Craske, 2010), obsessive-compulsive disorder (Skoog & Skoog, 1999), and eating disorders (Hudson et al., 2007). Long term follow-ups will provide information on new incidence and comorbidity, as well as information on recovery and remission rates, frequency of relapse, and recurrence for individuals meeting diagnostic criteria earlier in the study.

3.3 Questionnaire Issues

To make direct comparisons with data collected during the baseline and follow-up measures, a number of instrumentation issues should be considered, including the domains of interest, approaches to assessing mental disorders, the scope of the assessment, psychometric properties of the measures, instrument format, and similarities in the instruments used across time (from youth to adulthood and across the waves of data collection). The sections that follow explore these practical considerations.

3.3.1 Domains of Interest

To add a mental health longitudinal component to NSDUH, instruments would be needed that yield information about the prevalence and incidence of mental disorders, the course of mental disorders, risks factors for developing mental disorders, the services used by those who have mental disorders, and the effect those services have on the course of the disorder(s). Areas of particular importance to assess would be

- Mental disorders;

- Comorbid substance use disorders;
- Mental distress;
- Impairment across various areas of functioning (e.g., employment, psychiatric hospitalizations, interpersonal conflict);
- Course of mental disorders;
- Risk factors for the development of mental disorders;
- Mental health service utilization and barriers to treatment; and
- Perceived improvement from treatment of mental disorders.

The cost assumptions for each design scenario presented in Section 4 include all of the domains of interest above with the exception of risk factors for the development of mental disorders, mental health service utilization and barriers to treatment, and perceived improvement from treatment of mental disorders. These items could be included in the mental health longitudinal study under the assumed administration time, but other items would need to be removed, such as collecting data on low prevalence disorders.

3.3.2 Dimensional and Categorical Approaches to Assessing Mental Disorders

Combined with earlier international studies in psychiatric epidemiology, the Epidemiologic Catchment Area study, National Comorbidity Study, and related surveys demonstrated that mental disorders were highly prevalent in the general population. These studies have also changed how views on mental health and psychopathology. For example, most mental disorders begin far earlier in life than was previously believed. In addition, findings such as the frequency of comorbidity among the major diagnostic entities have led to systematic studies of their boundaries and renewed questions about categorical versus dimensional approaches to classification.

Although both clinical and research practices in mental health assessment are interested in diagnosis, the focus of clinical work is the individual's profile (dimensional approach), which allows for case conceptualization and treatment implications. In contrast, research practices are often more interested in whether or not the participant meets diagnostic criteria (categorical approach) for purposes of calculating prevalence rates of disorders and determining whether inclusion criteria are met.

3.3.3 Scope of the Assessment

There is also the issue of depth versus breadth: Depth affords more thorough assessment and fleshing out of specific symptoms, but for a smaller number of disorders. Breadth affords a thumbnail sketch of the key symptoms across a greater number of disorders. Focusing on depth or breadth also has implications for hiring and training interviewers: More depth will require masters- and doctoral-level clinicians with experience administering semi-structured diagnostic interviews (so they can exercise clinical judgment and probe as needed). If the focus is on breadth, then interviewers with little to no experience administering structured or semi-structured interviews would be suitable for administering more structured instruments (where all follow-up probes are incorporated).

3.3.4 Psychometric Properties

Decisions about instrumentation are heavily weighted on the reliability and validity of the candidate measures. In general, the weaknesses of questionnaires and other self-administered measures are that they tend to have inferior psychometric properties for the purposes of diagnosing mental disorders. Abbreviated instruments tend to have lower reliability coefficients than longer scales. For example, application of the Spearman-Brown prophecy formula (Crocker & Algina, 1986) suggests that even an adequate stability coefficient of .8 would be significantly reduced (to .65 or less) for a briefer version an instrument. Consequently, use of the short forms is not recommended as the primary source of diagnostic information (as part of the clinical follow up) (*The tenth mental measurements yearbook*, 1991).

3.3.5 Instrument Format

Some self-administered instruments can be used as a screening measure, a differential diagnosis measure, or a confirmatory measure. As a screening measure, self-administered instruments are used to detect individuals who may or may not need further (more costly) assessment. As a differential diagnosis measure, self-administered instruments are used to increase the accuracy of assessments that indicate the presence of two or more closely-related disorders. As a confirmatory measure, self-administered instruments are used to increase the accuracy of assessments that indicate the presence of one or more disorders. Therefore, the term "screening instrument" refers to how a measure is used, not to any intrinsic property of the measure itself. For example, the PDSQ is a questionnaire designed to measure psychiatric disorder symptomatology and can be used as a screening measure, differential diagnosis measure, or confirmatory measure.

Measures designed to screen for the DSM disorders are often used in conjunction with the structured or semi-structured diagnostic interview, allowing researchers and clinicians to determine which disorder(s) are indicated for a more in-depth diagnostic follow-up assessment; however, if there is only enough time to conduct a brief assessment, then some screening measures can also be used to extrapolate from self-report measures to clinical diagnosis, as long as there is research to support the validity of this approach. For example, based on cutoffs resulting in a sensitivity of 90 percent, the mean negative predictive value of the PDSQ subscales is 97 percent, meaning that 97 percent of the negative cases were correctly classified. Similarly, abbreviated diagnostic interviews can also be reasonable substitutes for a full diagnostic assessment. That said, some screening instruments are truly designed as "screeners" to alert the interviewer to areas that need to be followed up in an interview and do not work well as a stand-alone psychometric instrument. For example, the SCID-II screener as a stand-alone measure of Axis II diagnoses would not be advisable without following up with the SCID-II interview.

The use of a screening instrument is based on the research agenda and goals of the study—namely, whether the primary source of error should come from false positives or false negatives. Too many false positives may lead to the overestimation of the prevalence rates while too many false negatives may underestimate prevalence rates. For a true screening, the purpose is to reduce the number of people who need further (more costly) assessment by eliminating anyone who has no chance of being called a "case" upon further assessment. Because the concern here is false negatives, a cutoff score that is selected is low enough to minimize false

negatives but high enough to reduce the number of people who need further assessment. For differential diagnosis, there is an equal concern about false positives and false negatives, so a moderate cutoff is used to minimize both kinds of errors and provide the greatest percentage of correct decisions. For confirming a diagnosis, the concern is false positives; therefore, a cutoff score is selected to be high enough to minimize false positives but low enough to identify true (positive) cases. That said, almost any instrument that yields a dimensional score (e.g., PDSQ) can be used for screening, differential diagnosis, or confirming a diagnosis just by adjusting the cutoff score and therefore would be a suitable substitute for a diagnostic interview at the 6-month follow-up interviews. An abbreviated SCID or Composite International Diagnostic Interview (CIDI) might also be suitable for the 6-month follow-up.

Screening and structured measures are less expensive to administer than semi-structured diagnostic interviews because they take less time and can be administered by interviewers with little to no expertise. Another instrument issue for consideration is that copyrighted materials can have fees associated with their use.

3.3.6 Instrument Comparability

As a longitudinal study, it will be essential for there to be agreement in the measures used to assess youths and adults, as well as across the waves of data collection. Optimally, instruments for the youth and adult samples would be very similar to provide assessment continuity and comparability of the results across time. For example, the CIDI could be used for assessing youths and adults at baseline and follow-ups. However, the instrument version and disorders would still be different between the youths and adults.

The DSM-IV recognizes that some disorders are prominent mainly during childhood (e.g., separation anxiety) and other disorders are infrequent in childhood (e.g., schizophrenia). Moreover, disorders may exist among children and adults but present differently in each. For instance, there is growing evidence that attention deficit disorder continues into adulthood with a decrease in hyperactive symptoms and a continuance of inattention symptoms (Millstein, Wilens, Biederman, & Spencer, 1997).

These changes have necessitated the development of different assessments for mental illness in children and adults. This presents a challenge for researchers interested in the continuity of disorders from childhood to adulthood (individual trends) because changes in diagnosis or symptom profiles may be due to changes in the individual or changes from the measure.

This is a limitation to any longitudinal study. One method of addressing this limitation would be to conduct a "dual coding study." This would involve simultaneously administering the youth and adult measures at the assessment in which youths were 18 years old to discern what the individual's measures would have been if the youth assessments were continued and compare that to the estimates obtained in the adult measure. Unfortunately, this doubles the response burden on participants for that assessment, likely increasing loss to follow-up, and therefore may not be realistic. Another consideration would be to administer each instrument to a split-sample of respondents. Examples of studies that administer a split-sample to 18 year olds to assess the comparability of the youth questionnaire and the adult questionnaire were not readily available in

the literature. Findings about the comparability of the SCID and the K-SADS or other diagnostic interviews and instruments were also scarce in the literature. Conducting a split sample could assess this comparability and contribute to this literature.

3.3.7 Candidate Instruments

Based on the aforementioned considerations, a number of instruments could be well-suited for an adult and child longitudinal mental health study. This section describes potential measures for estimating incidence and prevalence and for assessing course of disorders.

3.3.7.1 Adult Mental Health Instruments

Table 2 provides a summary of the most widely used instruments to assess adult mental health. These instruments are further described below.

Diagnostic Instruments. Structured or semi-structured clinical interviews allow researchers and clinicians to conduct a more in-depth diagnostic assessment of the DSM disorders. This section describes three measures used for these purposes: the Structured Clinical Interview for DSM-IV-TR Axis I Disorders Non-Patient Edition (SCID-I/NP), the CIDI, and the Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV).

The Structured Clinical Interview for DSM-IV-TR Axis I Disorders Non-patient Edition (SCID-I/NP) (First et al., 2002) is considered the "gold standard" of diagnostic assessment and is the clinical interview measure used in the MHSS calibration study. The SCID-I/NP is a semi-structured interview that has been widely used in clinical calibration studies such as the NCS-R (Kessler et al., 2004), the National Survey of American Life (Jackson et al., 2004), and the NSDUH substance-use disorders reappraisal study (Jordan et al., 2008). It has demonstrated good reliability (Segal et al., 1995; Zanarini & Frankenburg, 2001; Zanarini et al., 2000) and validity (Fennig et al., 1994; Kranzler et al., 1996; Ramirez Basco et al., 2000; Shear et al., 2000; Steiner, Tebes, Sledge, & Walker, 1995). In addition to the diagnostic modules, SCID includes four other modules: (1) an open-ended overview module designed to elicit information about the respondent's diagnostic and treatment history and current status in a way that establishes some level of rapport between the interviewer and the respondent, (2) a screener module containing questions for several of the anxiety disorders and eating disorders, (3) a module containing the DSM-IV Axis V Global Assessment of Functioning (GAF) Scale (a clinical interviewer [CI] rating of the respondent's period of worst psychological, social, and occupational functioning during the past year), and (4) a module for documenting the CI's impressions of the interview situation, including ratings of the respondent's level of privacy, cooperation, and comprehension, as well as the overall validity of the interview data (any interview deemed by the CI or clinical supervision team to be of questionable validity was discarded). The SCID is available in 12-month and lifetime versions. It is currently administered using a paper-and pencil instrument, but there is an electronic version of the SCID under development. The SCID must be administered by trained clinical interviewers.

Table 2. Adult Mental Health Instruments

Instrument	Format	Length	Diagnostic Criteria	Diagnostic Coverage	Diagnostic Timeframe	Administered by	Studies Using this Instrument
SCID (Research Version)	Semi-structured interview	15–180 min., 60 min. average	DSM-IV	Major depressive episode Manic Episode Hypomanic episode Dysthymic disorder (current only) Mood disorder due to a general medical condition (GMC) Substance-induced mood disorder Panic disorder with agoraphobia Panic disorder without agoraphobia Agoraphobia without history of panic disorder Social phobia Specific phobia Obsessive-compulsive disorder Posttraumatic stress disorder Generalized anxiety disorder (current only) Anxiety disorder due to a GMC Substance-induced anxiety disorder Anxiety disorder NOS Schizophrenia Schizophreniform disorder Schizoaffective disorder Delusional disorder Brief psychotic disorder Psychotic disorder due to a GMC Substance-induced psychotic disorder Psychotic disorder NOS Bipolar I disorder Bipolar II disorder Other bipolar disorders Major depressive disorder Depressive disorder NOS Substance use disorders Somatization disorder (current only)	Lifetime for most disorders Current for all disorders	Clinicians or extensively trained lay interviewers	NSDUH – MHSS Collaborative Longitudinal Personality Disorders Study Suffolk County Mental Health Project The Netherlands Mental Health Survey and Incidence Study (NEMESIS) – Psychotic Disorder Module

(continued)

Table 2. Adult Mental Health Instruments (continued)

Instrument	Format	Length	Diagnostic Criteria	Diagnostic Coverage	Diagnostic Timeframe	Administered by	Studies Using this Instrument
SCID (Research Version) (continued)				Undifferentiated somatoform disorder (current only) Pain disorder (current only) Hypochondriasis (current only) Body dysmorphic disorder (current only) Anorexia nervosa Bulimia nervosa Binge eating disorder Adjustment disorder (current only) Acute stress disorder Minor depressive disorder Mixed anxiety depressive disorder Additionally, the SCID-II, measures personality disorders			
CIDI	Fully structured interview	Average 120 min.	DSM-IV and some ICD	DSM IV: Major depression Mania Anxiety panic disorder Specific phobia Agoraphobia Generalized anxiety disorder Post-traumatic stress disorder Obsessive-compulsive disorder Social phobia Substance abuse and dependence Childhood attention-deficit/hyperactivity disorder oppositional-defiant disorder Conduct disorder Separation anxiety disorder Other intermittent explosive disorder	Lifetime Current	Clinicians or trained lay interviewers	NCS-R, NCS-2 World Mental Health Survey Initiative NEMESIS

(continued)

Table 2. Adult Mental Health Instruments (continued)

Instrument	Format	Length	Diagnostic Criteria	Diagnostic Coverage	Diagnostic Timeframe	Administered by	Studies Using this Instrument
CIDI (continued)				Eating disorders Premenstrual disorder Non-affective psychoses screen Pathological gambling Neurasthenia Personality disorders screens			
AUDADIS-IV	Fully structured interview	Average 60 min.	DSM-IV	Major depression Dysthymia Mania Hypomania Generalized anxiety disorder Panic disorder (with and without agoraphobia) Social phobia Specific phobia Posttraumatic stress disorder Attention-deficit/hyperactivity disorder Pathological gambling Personality disorders Alcohol, tobacco, and drug consumption Substance use disorders	Current lifetime	Clinicians or trained lay interviewers	National Epidemiological Survey of Alcohol and Related Conditions (NESARC)

(continued)

Table 2. Adult Mental Health Instruments (continued)

Instrument	Format	Length	Diagnostic Criteria	Diagnostic Coverage	Diagnostic Timeframe	Administered by	Studies Using this Instrument
PDSQ	Fully structured self-report	20 min.	DSM-IV symptoms	Major depressive disorder Generalized anxiety disorder Panic disorder Posttraumatic stress disorder Alcohol abuse/dependence Drug abuse/dependence Psychosis Bulimia/binge-eating disorder somatization disorder Obsessive-compulsive disorder Social phobia Hypochondriasis Agoraphobia	Current	Self-administered	Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project
CIDI-Short Form	Fully-Structured interview	7–45 min.	DSM-IV and DSM-III-R	Major depression (DSM-IV) Generalized anxiety (DSM-IV) Specific phobia (DSM-IV) Social phobia (DSM-IV) Agoraphobia (DSM-IV) Panic attack (DSM-IV) Alcohol dependence and drug dependence (DSM-III-R)	Current and lifetime	Clinicians or trained lay interviewers	Health and Retirement Survey U.S. National Health Interview Survey

AUDADIS-IV = Alcohol Use Disorder and Associated Disabilities Interview Schedule, version IV; CIDI = Composite International Diagnostic Interview; MHSS = Mental Health Surveillance Study; NCS = National Comorbidity Study; NOS = not otherwise specified; NSDUH = National Survey on Drug Use and Health; PDSQ = Psychiatric Diagnostic Screening Questionnaire; SCID = Structured Clinical Interview for the DSM-IV.

The current MHSS assessment battery is designed to measure a number of domains that could also be useful for a mental health longitudinal component. Table 3 provides a summary of the mental health data currently gathered on the NSDUH and MHSS.

Table 3. Mental Health Surveillance Study Measures by Domain

Domain	Measure
Mental Disorders—Past Year and Lifetime Manic Episode; Past Year Dysthymic Disorder, Psychotic Symptoms, Posttraumatic Stress Disorder, Panic Disorder, Agoraphobia, Social Phobia, Specific Phobia, Obsessive-Compulsive Disorder, Generalized Anxiety Disorder, Anorexia, Bulimia, and Intermittent Explosive Disorder	Structured Clinical Interview for DSM-IV-TR (SCID)
Mental Disorders—Lifetime and Past Year Major Depressive Episode and Past Year Substance Use Disorders	SCID and the National Survey on Drug Use and Health (NSDUH)
Comorbid Substance Use Disorders	SCID
Mental Distress	NSDUH (K6) and the SCID
Impairment Across Various Areas of Functioning (e.g., Employment, Psychiatric Hospitalizations, Interpersonal Conflict)	NSDUH (World Health Organization Disability Assessment Schedule) and the SCID
Some Risk Factors for Developing Mental Disorders—Sex, Age, and Family History	NSDUH and SCID
Mental Health Service Utilization and Barriers to Treatment	NSDUH

A primary objective of a longitudinal study is to assess changes in mental disorder status that have occurred since the baseline assessment, namely changes that are associated with course of mental disorders. Although quite comprehensive, the current MHSS assessment battery lacks necessary measures for assessing course of mental disorders and risk factors for developing mental disorders, such as past-year disorders not already captured by the SCID, pre-existing mental disorders, and exposure to trauma. The current MHSS SCID interview takes about an hour to complete. Adding a lifetime component (necessary for incidence) and the full psychotic disorders assessment to the current MHSS SCID would require approximately 25 minutes, and a screener for assessing personality disorders (from the SCID-II) would require approximately 10 minutes.

The Composite International Diagnostic Interview (CIDI) is a comprehensive and fully standardized diagnostic interview designed for assessing mental disorders according to the definitions of the diagnostic criteria for research of ICD-10 and DSM-IV. The full-length instrument contains 276 symptom questions, many of which are coupled with probe questions to evaluate symptoms of depression, mania, panic disorder, specific phobia, social phobia, agoraphobia, generalized anxiety disorder, intermittent explosive disorder, substance use disorders, posttraumatic stress disorder, eating disorders, obsessive compulsive disorder, psychosis, gambling, and personality disorder screen. The CIDI also assesses disorders often identified in childhood, such as attention deficit/hyperactivity disorder, oppositional defiant disorder, and conduct disorder. The CIDI is available in lifetime and 12-month versions and in

both paper-and-pencil and computer-administered forms. The data from the CIDI are entered into standard data entry and scoring programs that give as output the diagnostic criteria satisfied.

The CIDI has two versions, the University of Michigan version (UM-CIDI) and the World Health Organization version (WMH-CIDI). The UM-CIDI was developed from the WMH-CIDI and contains changes to question wording to improve comprehension among U.S.-based participants as well as enhance memory probing for onset data. Estimates obtained using the UM-CIDI were compared to estimates using the WMH-CIDI. The UM-CIDI was then clinically appraised against the SCID. Compared to WMH-CIDI estimates, the UM-CIDI found higher estimates of agoraphobia and simple phobia as well as higher estimates of comorbidity. The comparison of the UM-CIDI to the SCID found that overall the SCID produces higher prevalence estimates of disorders but that these differences were not statistically significant in most cases. There was a significantly higher prevalence estimate of major depressive episodes and a significantly lower prevalence estimate of simple phobia in the UM-CIDI compared to the SCID. The CIDI can be administered by trained lay interviewers. It is available in lifetime and 12-month versions, and in both paper-and-pencil and computer-administered forms.

The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV) is a fully structured diagnostic interview that assesses lifetime, past 12 month, and current mental disorders according to DSM-IV criteria. The full-length instrument takes an average of 60 minutes to complete. The AUDADIS-IV assesses a number of DSM-IV-based disorders, including major depression, dysthymia, and mania; generalized anxiety disorder, panic disorder, social phobia, specific phobia, and posttraumatic stress disorder; attention-deficit/hyperactivity disorder; pathological gambling; several personality disorders; alcohol, tobacco, and drug consumption; and the substance use disorders, including nicotine dependence. The AUDADIS-IV also collects information on risk factors for mental illness, including family history of depression; acculturation and race-ethnic orientation; sexual orientation; experiencing discrimination; perceived stress; social support; adverse childhood experiences, mainly abuse and neglect; and intimate partner violence. The AUDADIS-IV differs from other clinical assessments in that it was designed to translate complex clinical constructs for better comprehension by lay persons, and it assesses a complete symptom profile from respondents instead of using skip-outs that end question strings once disorder criteria are reached (Grant, Dawson et al., 2003). The AUDADIS-IV has fair to excellent test-retest reliability in general population samples, which varies based on the diagnostic module (Grant, Dawson et al., 2003; Ruan et al., 2008). For example, inter-rater reliability coefficients are only fair to good for mood, anxiety, and personality disorders. The AUDADIS-IV is administered via computer (computer-assisted personal interviewing [CAPI]) by trained lay interviewers, with preprogrammed skip, logic, and consistency checks. Fewer disorders are assessed in the AUDADIS-IV than in the SCID or the CIDI.

Brief Instruments (for screening or six-month assessments). Measures designed to screen for the DSM disorders are primarily used in conjunction with the structured or semi-structured diagnostic interview, allowing researchers and clinicians to determine which disorder(s) are indicated for a more in-depth diagnostic assessment; however, screening measures can also be used to improve the efficiency of diagnostic evaluation by estimating symptom areas and/or mental disorders that could be examined in a follow-up assessment. These brief instruments could also be used for the proposed six-month assessments. However, a

recommended alternative for these brief assessments is to administer specific modules from the diagnostic interview administered at baseline or annual assessments. This would allow for better comparability between assessments.

Two measures are described below for these purposes: the Psychiatric Diagnostic Screening Questionnaire (PDSQ) and the Composite International Diagnostic Interview (CIDI) screens.

Psychiatric Diagnostic Screening Questionnaire (Zimmerman & Mattia, 2001) is a 126-item self-administered questionnaire designed to screen for the most common DSM-IV Axis I disorders encountered in an outpatient mental health setting. The PDSQ consists of 13 non-overlapping subscales that assess the symptoms of major depressive disorder, generalized anxiety disorder, panic disorder, posttraumatic stress disorder, alcohol abuse or dependence, drug abuse or dependence, psychosis, bulimia/binge-eating disorder, somatization disorder, obsessive-compulsive disorder, social phobia, hypochondriasis, and agoraphobia. Each item is rated as either yes or no. The response format is yes/no. The timeframe varies from 2 weeks (e.g., major depressive disorder, obsessive-compulsive disorder, panic) to 6 months (e.g., generalized anxiety disorder). Each scale contains 5–15 items (the average is 8). The items for each subscale are grouped together for easy scoring. Cutoff scores, critical items, and follow-up interview guides are provided for each disorder. There are no validity or treatment scales. The PDSQ requires 20 minutes to administer. The measure is copyrighted and will cost about $270 for each 500 administrations (1 manual + 50 test booklets + 500 answer sheets + computer software for printing follow-up interview "guides").

The Composite International Diagnostic Interview Short Form is a comprehensive and fully standardized diagnostic interview designed for assessing mental disorders according to the definitions of the Diagnostic Criteria for Research of ICD-10 and DSM-IV. The short version consists of 159 symptom questions, many of which are coupled with probe questions to evaluate symptom severity, as well as questions for assessing help-seeking behavior, psychosocial impairments, and other episode-related questions. There are no validity or treatment scales. The CIDI short form stem questions require about 45 minutes. The measure is copyrighted, but all of the materials are available on the website. The CIDI is available in lifetime and 12-month versions and in both paper-and-pencil and computer-administered forms. The data from the CIDI are entered into standard data entry and scoring programs that give as output the diagnostic criteria satisfied.

3.3.7.2 Youth Mental Health Instruments

Table 4 presents a summary of the most widely used instruments to assess youth mental health status. Information provided below describes candidate questionnaires to be considered as

Table 4. Instruments for Assessing Child and Adolescent Mental Health

Instrument	Format	Age Range	Time to Administer	Informants	Diagnostic Criteria	Diagnostic Coverage	Diagnostic Timeframe	Administered by	Studies Using this Instrument
K-SADS	Semi-structured interview (82-item screener also available)	6–18 years	1.5 hours child 1.5 hours parent	Child Parent	DSM-IV	Affective, anxiety, behavioral, easting, psychotic, substance abuse	Lifetime Past year (worst episode) Considered current if not symptom-free for past 2 months	Clinician	NCS-A—assessed K-SADS convergent validity with CIDI Oregon Adolescent Depression Project
DISC-IV	Structured interview	6–17 years	Mean 70 minute general population 90–120 min clinical population	Child Parent	DSM-IV ICD-10	Anxiety, mood, schizophrenia, behavior, substance use, misc. disorders; 30 disorders total	Past year Current (past 4 weeks) Age at onset	Trained lay interviewer	MECA NHANES
CAPA	Structured interview	9–18 years	1 hour parent 1 hour child	Child Parent	DSM-IV	Anxiety, mood, behavioral, substance, and antisocial personality	Lifetime Current (past 3 months) Age at onset	Trained lay interviewer	GSMS
NCS-A CIDI	Structured interview	13–17 years	2–3 hours (child)	Child Parent	DSM-IV	Mood, anxiety, disruptive behavior, substance use, eating, neurasthenia, and suicidality (19 total)	Lifetime Current	Trained lay interviewer	NCS-A
SDQ	25-item checklist; 5-item version Symptom checklist	4–17 years	4-5 minutes for 25-item checklist (child report); 1-2 minutes for 5-item version (parent report)	Child (11–17) Parent	DSM-IV	Broad problem scales—emotions, hyperactivity/inattention, conduct, and peer problems Also pro-social behaviors	Past 6 months	Self (parent, child)	NHIS British nationwide surveys of child MH
CIS	Structured questionnaire	7–17 years	5 minutes	Child Parent	N/A	Global impairment on four dimensions: interpersonal relationships, psychopathological domains, functioning in job or school, and use of leisure time	Not specified	Trained lay interviewers Self (in large population-based surveys)	MEPS

(continued)

Table 4. Instruments for Assessing Child and Adolescent Mental Health (continued)

Instrument	Format	Age Range	Time to Administer	Informants	Diagnostic Criteria	Diagnostic Coverage	Diagnostic Timeframe	Administered by	Studies Using this Instrument
CGAS	Single scale with a range of 1–100	4–16 years	No administration time because it is based on prior clinical assessment; 5 minutes to generate score	Clinician-trained lay interviewer	N/A	Unidimensional and global measure of social and psychiatric functioning	Not specified	Clinician-trained lay interviewer	

CAPA = Child and Adolescent Psychiatric Assessment; CGAS = Children's Global Assessment Scale; CIDI = Composite International Diagnostic Interview; CIS = Columbia Impairment Scale; DISC-IV = Diagnostic Interview Schedule for Children Version IV; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders IV; GSMS = Great Smoky Mountain Study; ICD-10 = International Classification of Diseases 10; K-SADS = Kiddie-Schedule of Affective Disorders and Schizophrenia; MECA = Methods for the Epidemiology of Child and Adolescent Mental Disorders; MEPS = Medical Expenditures Panel Survey; MH = mental health; NCS-A = National Comorbidity Survey Replication Adolescent Supplement; NHANES = National Health and Nutrition Examination Survey; NHIS = National Health Interview Survey; SDQ = Strength and Difficulties Questionnaire.

diagnostic interviews, assessments of functional impairment, and measures for use in sampling. Some operational issues in the administration of these questionnaires are also described. This section does not include a description of supplemental measures that might be considered to examine risk or protective factors related to mental disorders, including measures of mental health service receipt. The use of these measures is not included in the cost assumptions for the proposed design scenarios in Section 4. There are candidate measures to assess these constructs for youths which could be considered in the future.

Diagnostic Instruments. For assessing mental health diagnoses in youth, there are several candidate diagnostic interview instruments. These instruments have been selected as candidates because of their strong psychometric properties, use in large-scale psychiatric epidemiology studies, and comprehensive coverage equivalent to an instrument such as the SCID for adults. The most widely used diagnostic interviews for youths are the Kiddie-Schedule of Affective Disorders and Schizophrenia (K-SADS), the Diagnostic Interview Schedule for Children (DISC), and the Child and Adolescent Psychiatric Assessment (CAPA). Table 4 provides some of the basic parameters of these instruments. All of these diagnostic interviews recommend completing interviews from both parents and youth; however, specific study practices vary in their decision rules for collecting specific diagnostic information by parent and/or youth report.

Both the DISC and the CAPA are recommended for administration by trained lay interviewers. The K-SADS requires a clinically trained interviewer for its administration, similar to the current clinical follow-up in the NSDUH MHSS. Due to the proprietary nature of the instrument, the CAPA interview requires collaboration with the instrument developer (Duke University staff) for training, administration (done through the *e-measures* system necessary for CAPA administration), and scoring. The CAPA interview also requires administration via a tablet laptop computer in which the interviewer can "write" notes directly onto the computer screen. This option would be less desirable to the government due to the inability for the government to directly obtain and work with the data and the lack of transparency in scoring procedures (as all CAPA/Pre-School Age Psychiatric Assessment scoring algorithms are proprietary to Duke University).

A fourth option for assessing adolescent disorders is the CIDI. The National Comorbidity Survey Replication Adolescent Supplement (NCS-A) was built from the NCS-R sample to assess prevalence of mental health disorders in U.S. adolescents. The CIDI was adapted for use with youths aged 13–17 years (Merikangas et al., 2010). Four broad classes of disorder were measured: mood, anxiety, disruptive behavior, and substance use. Across these four categories and an "other" category, including some less common disorders, a total of 21 specific disorders were assessed. The CIDI was then administered to adolescents in face-to-face interviews using computer-assisted interviewing (CAI) techniques. The adolescent CIDI has shown concordance with the K-SADS for most major disorders (Kessler et al., 2009). As with the adult CIDI, the adolescent version includes lowered diagnostic thresholds for all disorders, so prevalence estimates are slightly higher than with clinical assessments.

In the NCS-A, a separate parent-report version of the adolescent CIDI was developed and administered to parents via paper-and pencil interviewing (PAPI). This mode was less preferable than face-to-face interviews, but parent self-report was necessary due to budget constraints.

Although the adolescent CIDI was adapted directly from the adult CIDI, it is not clear how the parent self-report was developed, and no published results have been identified on its reliability or validity. As such, the instrument developers would need to be engaged to determine whether and how to use the parent version of the adolescent CIDI to assess its added value to the adolescent instrument and possibly adapt it to a different mode of administration. Despite the relative unknowns inherent to using the parent instrument, the CIDI offers considerable benefits—most notably a matched instrument across youths and adults that would allow assessment of disorder continuity during the transition to adulthood. An additional benefit is the use of lay interviewers instead of clinical interviewers; moreover, the same pool of interviewers could be used for both the adult and adolescent studies. This would lead to considerable cost savings and was one of the primary justifications for the downward adaptation of the CIDI for the NCS-A (Kessler et al., 2009).

Administering a full diagnostic interview to youth will take 60 to 180 minutes of interview time to complete both the parent and youth interview. The CAPA requires an additional hour per interview for the interviewer to code the instrument after the respondent interview has been completed. These estimates are exclusive of any other measures that may be added to capture risk and protective factors and/or service use. The time burden would be smaller in non-clinical cases because fewer questions would be asked. Each instrument described can be conducted in person or over the phone, with proper training and distress protocols.

The typical large-scale administration for these diagnostic instruments is face-to-face in the field. For instance, this was the primary data collection method used for the NCS-A and the Great Smoky Mountain Study (GSMS). However, researchers have conducted psychiatric clinical interviews over the telephone for many years (Rohde, Lewinsohn, & Seeley, 1997; Sobin et al., 1993). In youth psychiatric assessment, this has become operational practice with hard-to-reach participants. For instance, Duke University investigators estimate that approximately 112 CAPA interviews in the Great Smoky Mountain study were collected with adolescents over the telephone with no adverse effects (Personal communication with Dr. Jane Costello, September 2010). Telephone administration has also become widely accepted as a way to conduct clinical re-appraisals. For instance, Kessler and colleagues conducted K-SADS interviews by telephone with 347 adolescent participants and their parents in the NCS-A as part of the clinical reappraisal progress (Kessler et al., 2009). They compared telephone K-SADS results to the field-based CIDI results. Kessler and colleagues concluded that there was strong individual-level concordance between the CIDI and K-SADS for most diagnoses. While CIDI estimates were higher than K-SADS estimates for specific phobia and oppositional defiant disorder, estimates of the prevalence of any disorder was only slightly higher in the CIDI than K-SADS.

Two studies have compared face-to-face and telephone administration of diagnostic interviews on child mental health measures. Lyneham and Rapee (2005) report on a comparison of diagnoses using the Anxiety Disorders Interview Schedule for Children for DSM-IV (ADIS-C-IV). Diagnoses from parent interviews over the telephone were compared with diagnoses from face-to-face administration using both the child and parent versions. They report good to high levels of agreement for a number of disorders including separation anxiety, generalized anxiety, social phobia, specific phobia, attention-deficit hyperactivity, oppositional defiant, and dysthymia. In addition, Paing, Weller, Dixon, and Weller (2010) report on a study comparing

telephone vs. face-to-face administration of the parent version of the Children's Interview for Psychiatric Syndromes (P-ChIPS), finding high levels of agreement for 21 syndromes.

So, while not nearly as widespread as in-person data collection method, the feasibility of telephone administration has been demonstrated with some of these diagnostic instruments. This type of administration may offer the ability to centralize data collection and closely supervise interviewers without the geographic restrictions required for face-to-face clinical assessments.

Brief Instruments (for screening or six-month follow-up assessments). For the 6-month follow-ups, disorders identified at baseline and annual waves will be tracked. In keeping with the methodology for the adult interview, individual CIDI modules could be used to follow up on specific disorders. The K-SADS has a screener for this purpose as well, but it requires clinical interviewers and is linked to the K-SADS long form; as such, the K-SADS screener would work best if the annual assessments were to use the K-SADS. The SDQ would provide a brief measure of broad symptom domains, but would be less disorder-specific than the CIDI modules, yielding less information on the path of any given disorder over time.

Functional Impairment. A complementary construct to assess when measuring mental health symptoms and diagnoses is functional impairment. The diagnostic interviews discussed in the previous section offer an assessment of impairment. For instance, the version of the CIDI used in the National Comorbidity Survey for adults and adolescents includes an expanded version of the Sheehan Disability Scale after each diagnostic module to assess overall level of impairment associated with a disorder. The CAPA includes an extensive impairment module used in the diagnostic algorithm to calculate the prevalence of mental disorders. The DISC incorporates a series of impairment questions after each diagnostic module. Also, the K-SADS has a Clinical Global Impressions Scale at the end of the instrument that is designed to assess impairment.

It may also be informative to assess impairment at the 6-month follow-up assessment. For this, it may be better to rely on the impairment questions that follow each diagnostic module of the CIDI. However, if a different measure is chosen for the baseline and annual assessments, a short impairment measure may need to be added to the selected follow-up screener. Child functional impairment is often measured separately from mental health symptoms or diagnoses. Two candidate instruments include the Columbia Impairment Scale (CIS; Bird et al., 1993) or the Children's Global Assessment Scale (CGAS; Shaffer et al., 1983).

The use of the CGAS does not appear ideal. Although it is based on an adaptation of the Global Assessment Scale (GAS) for adults (and used within the NSDUH MHSS), it has not been used within any of the large child epidemiological studies reviewed for this report. The CGAS also must be administered in the context of a diagnostic interview because it involves a rating provided based on results of a prior clinical interaction. The use of the CIS is recommended. It has been used in several mental health surveillance efforts, including the MEPS, has a short administration time, and has strong psychometric properties.

Measures for Use in Sampling. A short child mental health measure could be used to facilitate sample selection for the mental health longitudinal study. This type of measure would be administered during the main NSDUH interview. This instrument would serve a comparable

purpose to the K-6 and World Health Organization Disability Assessment Schedule (WHODAS) for the adult sample selection process. The inclusion of these items will increase the NSDUH survey administration time which may potentially adversely affect response rates and increase respondent burden. However, there are also a number of other existing variables on the NSDUH that can be used in the sampling process such as youth MDE and mental health treatment.

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) is the only brief children's mental health standardized measure currently used to assess children's mental health symptoms in two national population-based surveys, one in the United States (see Bourdon et al., 2005) and another in Britain. There are several versions of the SDQ—including the original 25-item instrument (for youth or parent report) and a 5-item version (designed specifically for parent report in the NHIS only). The 25-item screener includes five scales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior. The SDQ also has published norms for non-clinical samples in household populations. The 25-item checklist has been administered to parents and children. There is a 5-item "impact supplement" for this version of the instrument that assesses functional impairment. A 5-item version of the parent SDQ checklist was created for use within the U.S. National Health Interview Survey (NHIS) by taking the first item from each of the five scales within this measure. This version was developed in consultation with the instrument developer, but has not received psychometric testing outside of its use within the NHIS. This version is routinely given to parent participants of the NHIS and has a 1-minute administration time. The 5-item SDQ in the NHIS is supplemented by 1 "impact" item that assesses functional impairment. No comparable 5-item version has been developed for administration to youth.

There are other candidate screening measures used in national surveillance efforts, but none demonstrate the psychometric rigor and national comparison data available with the SDQ. For instance, the Youth Risk Behavior Surveillance System (YRBS) uses project-developed items to assess depression along with suicidal thoughts and behaviors, but not an indicator of general mental health status. The Medical Expenditures Panel Survey (MEPS) does include the Columbia Impairment Scale (CIS) and a 36-item Short Form Health Survey (SF36); however, the CIS does not measure symptoms, and the SF36 does not target a specific age or disease group. Finally, while the National Health and Nutrition Examination Survey (NHANES) included a mental health follow-up study from 2001 to 2004, the core NHANES survey does not include any ongoing measure of children's mental health status.

3.3.8 Statistics of Test Performance

The categorical approach in the development of clinical instruments in psychology has involved a focus on the ability of a particular measure to distinguish between members and non-members of a diagnostic group. Often these qualitative distinctions are made on the basis of quantitative scales—a cutoff is derived for identification purposes, which varies in diagnostic efficiency as a function of the setting in which the instrument is used (e.g., inpatient, outpatient).

Studies that have a small sample or a convenience sample often perform more comprehensive, specific personality assessments, whereas larger scale or epidemiological studies will sacrifice specificity for sensitivity so that a more representative sample can be used.

- Sensitivity vs. Specificity
  - Sensitivity refers to a test's ability to correctly identify individuals with the disorder
  - Specificity refers to the test's ability to identify persons without the disorder
- Positive vs. Negative Predictive Value
  - Positive predictive value refers to the probability that an individual has the disorder given that the test has identified the person as positive (percentage of true positives)
  - Negative predictive value refers to the probably that an individual does not have the disorder given that the test has identified the person as negative (percentage of true negatives)

Sensitivity, specificity, and positive and negative predictive values are a function of the cutoff point used to distinguish cases from non-cases and of the disorder's prevalence, and they are all are related to each other. Five axioms characterize these relationships:

1. Lowering a test's cutoff score to identify cases increases the test's sensitivity and decreases the specificity.
2. Raising the test's cutoff score to identify cases decreases the test's sensitivity and increases the specificity.
3. At constant sensitivity and specificity, the test's positive predictive value is higher in samples where the disorder is more prevalent (because, by chance alone, a test is more likely to predict a disorder that occurs more frequently in the sample distribution).
4. At constant sensitivity and specificity, a test's negative predictive value is higher in samples where the disorder is less prevalent (because, by chance alone, a test is less likely to predict a disorder that occurs less frequently in the sample distribution).

Because the mental health longitudinal study is an epidemiological study, sensitivity may supersede specificity, particularly for those disorders that are of secondary interest (e.g., personality disorders) and have relatively low prevalence rates. Questionnaires and other brief self-administered scales (e.g., the SCID-II self-report, Psychiatric Diagnostic Screening Questionnaire [PDSQ]) are typically designed to have high sensitivity to serve as screening measures that alert the interviewer to personality traits and disorders that need further assessment—that is, they have high sensitivity and relatively low specificity.

3.4 Sample Issues

This section provides an evaluation of sample issues for consideration, including the overall sampling strategy for selecting the mental health longitudinal sample, prevalence estimates for the mental disorders of interest, and design effects used to inform the sample size.

3.4.1 Sampling Strategy

The overarching goal of the NSDUH MHSS is to develop models that can be used to provide accurate estimates of the prevalence of serious mental illness (SMI) among adults aged 18 or older at national and state levels. However, the MHSS data have potential for a variety of important analyses beyond this primary purpose. Methods for estimating other categories of mental illness (e.g., "mild," "moderate," or "any" mental illness) have been developed. Further study of SMI and other mental illness models can be done with the MHSS dataset (e.g., to evaluate the current models' performance or to attempt to identify improved models). The MHSS data could also be used to generate estimates of specific disorders.

The MHSS respondents are recruited by NSDUH field interviewers to participate in this follow-up survey and then contacted via telephone by clinical interviewers for data collection in an expanded mental health module. The clinical interview is conducted using the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I) to assess mental illness (First, Spitzer, Gibbon, & Williams, 2002).

A sampling strategy could be used for the mental health longitudinal study comparable to the one used for the MHSS. The current MHSS selection algorithm oversamples respondents who are likely to be diagnosed with mental illness according to their K6/WHODAS scores. By oversampling these persons, the selection algorithm increases the proportion of positive SMI cases from approximately 5 percent (national estimate) to about 18 percent. Similarly, the proportion of positive any mental illness (AMI) cases increases from 20 percent to approximately 48 percent. The current MHSS selection algorithm also reverses the oversampling of younger age groups in the NSDUH.

As with the MHSS, respondents aged 18 or older would be selected for the mental health longitudinal study based on their K6/WHODAS scores in the main NSDUH and oversampled to meet the number of cases needed to generate national estimates of mental health disorder prevalence and incidence. Unlike the planning of the 2008 MHSS, there is now an advantage of over 3 years' experience implementing the MHSS to assist in designing the longitudinal study.

For respondents aged 12 to 17, given the burdens associated with the inclusion of an adolescent self-reported mental health measure, such as the SDQ, other currently collected variables such as MDE or mental health treatment would be used in the youth sampling algorithms. As with the K6/WHODAS scores among adults, the adolescent mental health measure would be used to select adolescents for the longitudinal component. Selected adolescents and their parents would complete a baseline diagnostic interview comparable to the adult diagnostic interview and would be followed for the longitudinal component.

3.4.2 Sample Design Advantages and Disadvantages

This sampling strategy may provide some benefits. First, the mental health longitudinal recruitment procedures would be similar to those approved by the IRB for the MHSS. This means that respondents would not be notified of the longitudinal nature of the study until the end of the main study interview. If OMB and the IRB would approve it, it would minimize the impact on main study response rates. However, it is unclear whether an IRB and OMB would

approve such a significant change from the MHSS processes. Second, using NSDUH data to oversample high-risk respondents for the MHSS and longitudinal component results in smaller sample size requirements, thus reducing labor and associated costs. Oversampling of persons with mental disorders is less critical in a longitudinal study such as this compared to the MHSS, because the MHSS was focused on developing models predicting SMI and therefore there was a need for an adequate number of cases with SMI. In a longitudinal study, where incidence is a key outcome measure, the sampling of persons with current mental disorders is less crucial. This design also enables the longitudinal sample to be recruited in person by field interviewers. However, this design also assumes the absence of any refusal conversion by field staff.

As discussed before, a brief mental health measure, such as the SDQ, could be added to the main NSDUH for follow-up sampling, which would increase the length of the main instrument for youths thereby potentially increasing nonresponse rates on the main study and increasing respondent burden. For example, depending on the version of the SDQ selected for inclusion in NSDUH, an average of 1 to 5 minutes would be added. The 25-item youth self-report SDQ would add 4 to 5 minutes of administration time, and a 5-item version would likely add 1 minute (comparable to the NHIS administration time for parents). Similar to the MHSS, there would also be some additional burden at the end of the main interview to recruit the respondent for the longitudinal component. Since increasing the length of the adolescent interview is not a desirable option, existing NSDUH youth mental health questions such as those on mental health treatment, contemplating suicide and major depressive episode (MDE) could be used for follow-up sampling. A recent analysis of MHSS data showed that seeking treatment for a mental disorder is a good predictor of SMI in adults.

Another sampling option would be to select the longitudinal sample at the NSDUH screening stage. However, there are several drawbacks to this approach. To have no impact on the main NSDUH, the screening instrument would need to remain the same and thus include only demographic data. Because nothing would be known about the respondent's mental history, a very large sample size would be required to estimate prevalence, incidence, and course with precision. Reducing this sample size would require a mental health screener that would likely be considered intrusive by respondents at this point early in the process and thus could have an impact on response to the NSDUH main study. Also, the screening respondent would be required to answer these mental health screener items about all household members, and the data collected may not be accurate. Finally, because the interviewer would be collecting sensitive data, this addition would also require a change to the informed consent process for the screening.

3.4.3 Sample Size

Estimates of prevalence and incidence from a literature review were used to inform the sample size for the NSDUH mental health longitudinal study. Estimates were drawn from the Outcomes of Depression International Network (ODIN), the NCS-R, the Health 2000 Study, a school-based study of Jakobstad, Finland, a published major depressive disorder meta-analysis, and the World Mental Health Survey Initiative. While NESARC estimates were also examined, estimates were not included from this study. Most publications of NESARC present 12-month prevalence and not lifetime estimates. Also, most of the publications present specific disorders and not disorder groups. Finally, the NESARC did not assess several disorders or else it assessed

only one disorder in a disorder group. For youths, estimates from GSMS and NHANES were examined.

Tables 5 and 6 provide summaries of the prevalence rates for adult and youth-specific disorders, respectively. For the purpose of evaluating sample size, the disorders were grouped into high-, medium-, and low-prevalence disorders. Then, an average prevalence was selected to represent each group. With the exception of bipolar and major depressive disorders, disorder groups were evaluated and not specific disorders. The availability of incidence data was very limited in the literature. For the few disorders with incidence estimates identified in the literature, the ranges were between 1 and 4 percent.

Table 5. Adult Disorders by Prevalence Group

Prevalence Group	Disorder or Disorder Group	Lifetime Prevalence	Data Sources
Low prevalence (0.3%–3.9%)	Adjustment disorders	0.3% (current prevalence aged 18 to 64); 2.3% (current prevalence aged 65 to 96)	ODIN
	Bipolar (I or II Combined)	3.9% (aged 18 or older)	NCS-R
	Psychotic disorders	2.29% (aged 30 or older)	The Health 2000 Study
Medium prevalence (7.1%–19.2%)	Eating disorders	7.1% (15-year-old females)	School-based study of Jakobstad, Finland
	Major depressive disorder	6.7% to 19.2% (15 or older and 18 or older)	Published meta-analysis and World Mental Health Survey Initiative
High prevalence (20.8%–28.8%)	Anxiety disorders	28.8% (aged 18 or older)	NCS-R
	Impulse-control disorders	24.8% (aged 18 or older)	NCS-R
	General mood disorders	20.8% (aged 18 or older)	NCS-R

ODIN = Outcomes of Depression International Network; NCS-R = The National Comorbidity Survey–Replication.

Table 6. Youth-Specific Disorders by Prevalence Group

Prevalence Group	Disorder Group	12-month Prevalence	Study Source
Low prevalence (2.1%–8.6%)	Attention deficit and disruptive behavior disorders	2.1%–8.6% (aged 8 to 15)	NHANES
	Mood disorders	3.7% (aged 8 to 15)	NHANES
High prevalence (13.1%–20.3%)	Any mental health disorder	13.1% and 20.3% (aged 8 to 15 and 9 to 13; former includes only a subset of major mood and behavior disorders, latter is 3-month prevalence)	NHANES and GSMS, respectively

GSMS = The Great Smoky Mountain Study; NHANES = National Health and Nutrition Examination Survey.

Assumptions about design effect were also incorporated into the sample size evaluation. Since there are no comparable studies for youth, the design effects from the adult MHSS were used for this effort. For several disorders, including SMI and AMI, Table 7 shows prevalence estimates, standard errors, and the corresponding design effects from the 2010 MHSS. The design effects are overall design effects; that is, they take into account clustering at the main study level as well as the MHSS sample design. Any changes in actual design effects would affect the effective sample size for the longitudinal survey.

The design effects for SMI and AMI are 0.8 and 1.7, respectively. The majority of the disorder-specific design effects fall within this range or are better. Thus, using this range of design effects and low, medium, and high prevalence rates, the precision that could be obtained from various sample sizes was evaluated for estimating incidence and remission. Tables 8 through 10 evaluate several sample sizes for the adult cohort, and Tables 11 and 12 evaluate sample sizes for the youth cohort. Because at-risk persons will be oversampled, it is expected that the actual numbers of persons with a diagnosis at time 1 (n2) will be larger than those shown in Table 8 to 12. Because being judged at-risk *ex ante* is not likely to be strongly correlated with the probability of remission, the variability of selection probabilities among person with the diagnosis at time 1 reduces the effective sample sizes. The numbers displayed in the tables serve as a useful starting point for the calculations, with the two design effects (0.8 and 1.7) adding a reasonable range to the results.

The estimation of prevalence rates generally require much smaller sample sizes. However, when considering course (which includes incidence measures), because a much smaller percentage of the population will be diagnosed at time 1, the ability to estimate remission with adequate precision requires a much larger sample size. In fact, a sample size of 10,000 in each of the youth and adult cohorts or 20,000 total persons would be sufficient to assess the course of high-prevalence disorders, but insufficient for assessing low-prevalence disorders and questionable for medium-prevalence disorders. Overall, a sample size of 5,000 in each cohort (or a total of 10,000 persons) would satisfy the objectives of the study. However, in order to be able to assess remission/course for subgroups (e.g., gender or race-ethnicity), the sample size would need to be doubled to 20,000.

One final sampling consideration is that different sample sizes for the youth and adult cohorts may be appropriate, depending on the goals of the study. For example, 5,000 youths and 10,000 adults may be sampled in order to be able to study gender-specific diagnoses in the latter group. The sample sizes in Tables 8 through 12 represent the sample size needed to study long-term course (i.e., the sample size needed at the end of the longitudinal study). A larger initial sample size will be needed to account for response rates and expected attrition over time.[2]

[2] The initial sample size is sufficient to yield the target sample size at the end of the study, and no sample refreshment is needed.

Table 7. Design Effects from the 2010 Mental Health Surveillance Study SCID Data (18 or older)

Variable[a]	Estimate (%)	Standard Error	Design Effect
Mood Disorders			
Past Year Major Depressive Episode (MDE)[b,c]	8.74	1.25	1.01
Lifetime MDE[d]	24.35	2.70	2.04
Past Year Manic Episode[b]	0.53	0.29	0.84
Lifetime Manic Episode	0.72	0.31	0.68
Dysthymic Disorder[b]	2.27	0.81	1.53
Lifetime Bipolar Disorder	0.74	0.31	0.68
Lifetime MDE or Lifetime Manic Episode	24.41	2.70	2.04
Past Year Any Mood Disorder	9.52	1.29	0.99
Any Mood Disorder	24.83	2.71	2.03
Past Year Psychotic Disorders			
Psychotic Screen[b]	0.24	0.13	0.38
Past Year Anxiety Disorders			
Posttraumatic Stress Disorder[b]	0.51	0.22	0.48
Panic Disorder with and without Agoraphobia[b]	0.86	0.32	0.61
Agoraphobia without History of Panic Disorder[b]	0.03	0.03	0.18
Social Phobia[b]	0.71	0.48	1.68
Specific Phobia[b]	0.93	0.63	2.23
Obsessive Compulsive Disorder[b]	0.23	0.12	0.32
Generalized Anxiety Disorder[b]	1.49	0.35	0.43
Any Anxiety Disorder	4.15	0.90	1.05
Past Year Eating Disorders			
Anorexia Nervosa[b]	0.03	0.03	0.18
Bulimia Nervosa[b]	0.04	0.04	0.18
Any Eating Disorder	0.07	0.05	0.18
Past Year Impulse Control Disorders			
Intermittent Explosive Disorder[b]	0.08	0.05	0.17
Past Year Substance Use Disorders (SUD)			
Alcohol Abuse[e]	4.88	1.12	1.40
Alcohol Dependence[e]	3.07	1.19	2.47
Non-Alcohol Substance Abuse[e]	0.78	0.39	1.01
Non-Alcohol Substance Dependence[e]	1.06	0.55	1.51
Any Substance Use Disorder	8.61	1.73	1.96
Past Year Adjustment Disorders			
Adjustment Disorder[b]	7.67	1.56	1.77
Any Disorder (including SUD and Adjustment)			
1 Disorder	18.58	2.53	2.18
2 Disorders	3.15	0.81	1.10
3+ Disorders	2.54	0.84	1.46
Any Disorder	24.27	2.60	1.89
Serious Mental Illness			
SMI	4.01	0.77	0.80
Any Mental Illness			
AMI	18.72	2.24	1.69

(continued)

Table 7. Design Effects from the 2010 Mental Health Surveillance Study SCID Data (18 or older) (continued)

Variable[a]	Estimate (%)	Standard Error	Design Effect
Serious Psychological Distress			
SPD	14.03	1.30	0.72

AMI = any mental illness; MHSS = Mental Health Surveillance Study; SCID = Structured Clinical Interview for the DSM-IV; SMI = serious mental illness; SPD = serious psychological distress.

[a] Missing values are treated as 0 for all disorder variables.

[b] Disorder was included in the estimation of gold-standard serious mental illness and any mental illness for the 2008 NSDUH MHSS's SCID sample.

[c] Estimate includes depressive episodes resulting from bereavement, general medical condition, or substance use.

[d] Estimate excludes depressive episodes resulting from bereavement, general medical condition, or substance use.

[e] Consistent with the DSM-IV diagnostic criteria, if an individual has both abuse and dependence for a specific substance, he or she is counted in the estimates for dependence on that substance.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2010.

Table 8. Adult Cohort: Low Prevalence Disorders (Adjustment Disorders, Bipolar, Psychotic Disorders), *p* = .03

Total Number (n)	Well at Time 1 (n1)	Probability (p) of Diagnosis (p1)	Standard Error (SE) (p1), deff = 0.8	SE (p1), deff = 1.7	Diagnosis at Time 1 (n2)	Probability of Remission (p2)	SE (p2), deff = 0.8	SE (p2), deff = 1.7
500	485	0.01	0.004	0.006	15	0.01	0.023	0.033
		0.02	0.006	0.008		0.03	0.039	0.057
		0.03	0.007	0.010		0.06	0.055	0.080
		0.04	0.008	0.012		0.09	0.066	0.096
		0.05	0.009	0.013		0.12	0.075	0.109
		0.06	0.010	0.014		0.15	0.082	0.120
		0.07	0.010	0.015		0.18	0.089	0.129
		0.08	0.011	0.016		0.21	0.094	0.137
		0.09	0.012	0.017		0.25	0.100	0.146
		0.10	0.012	0.018		0.30	0.106	0.154
1,000	970	0.01	0.003	0.004	30	0.01	0.016	0.024
		0.02	0.004	0.006		0.03	0.028	0.041
		0.03	0.005	0.007		0.06	0.039	0.057
		0.04	0.006	0.008		0.09	0.047	0.068
		0.05	0.006	0.009		0.12	0.053	0.077
		0.06	0.007	0.010		0.15	0.058	0.085
		0.07	0.007	0.011		0.18	0.063	0.091
		0.08	0.008	0.011		0.21	0.067	0.097
		0.09	0.008	0.012		0.25	0.071	0.103
		0.10	0.009	0.013		0.30	0.075	0.109
5,000	4,850	0.01	0.001	0.002	150	0.01	0.007	0.011
		0.02	0.002	0.003		0.03	0.012	0.018
		0.03	0.002	0.003		0.06	0.017	0.025
		0.04	0.003	0.004		0.09	0.021	0.030
		0.05	0.003	0.004		0.12	0.024	0.035
		0.06	0.003	0.004		0.15	0.026	0.038
		0.07	0.003	0.005		0.18	0.028	0.041
		0.08	0.003	0.005		0.21	0.030	0.043
		0.09	0.004	0.005		0.25	0.032	0.046
		0.10	0.004	0.006		0.30	0.033	0.049
10,000	9,700	0.01	0.001	0.001	300	0.01	0.005	0.007
		0.02	0.001	0.002		0.03	0.009	0.013
		0.03	0.002	0.002		0.06	0.012	0.018
		0.04	0.002	0.003		0.09	0.015	0.022
		0.05	0.002	0.003		0.12	0.017	0.024
		0.06	0.002	0.003		0.15	0.018	0.027
		0.07	0.002	0.003		0.18	0.020	0.029
		0.08	0.002	0.004		0.21	0.021	0.031
		0.09	0.003	0.004		0.25	0.022	0.033
		0.10	0.003	0.004		0.30	0.024	0.034

Table 9. Adult Cohort: Medium Prevalence Disorders (Eating Disorders, Major Depressive Disorder), *p* = .07

Total Number (n)	Well at Time 1 (n1)	Probability (p) of Diagnosis (p1)	Standard Error (SE) (p1), deff = 0.8	SE (p1), deff = 1.7	Diagnosis at Time 1 (n2)	Probability of Remission (p2)	SE (p2), deff = 0.8	SE (p2), deff = 1.7
500	465	0.01	0.004	0.006	35	0.01	0.015	0.022
		0.02	0.006	0.008		0.03	0.026	0.038
		0.03	0.007	0.010		0.06	0.036	0.052
		0.04	0.008	0.012		0.09	0.043	0.063
		0.05	0.009	0.013		0.12	0.049	0.072
		0.06	0.010	0.014		0.15	0.054	0.079
		0.07	0.011	0.015		0.18	0.058	0.085
		0.08	0.011	0.016		0.21	0.062	0.090
		0.09	0.012	0.017		0.25	0.065	0.095
		0.10	0.012	0.018		0.30	0.069	0.101
1,000	930	0.01	0.003	0.004	70	0.01	0.011	0.016
		0.02	0.004	0.006		0.03	0.018	0.027
		0.03	0.005	0.007		0.06	0.025	0.037
		0.04	0.006	0.008		0.09	0.031	0.045
		0.05	0.006	0.009		0.12	0.035	0.051
		0.06	0.007	0.010		0.15	0.038	0.056
		0.07	0.007	0.011		0.18	0.041	0.060
		0.08	0.008	0.012		0.21	0.044	0.063
		0.09	0.008	0.012		0.25	0.046	0.067
		0.10	0.009	0.013		0.30	0.049	0.071
5,000	4,650	0.01	0.001	0.002	350	0.01	0.005	0.007
		0.02	0.002	0.003		0.03	0.008	0.012
		0.03	0.002	0.003		0.06	0.011	0.017
		0.04	0.003	0.004		0.09	0.014	0.020
		0.05	0.003	0.004		0.12	0.016	0.023
		0.06	0.003	0.005		0.15	0.017	0.025
		0.07	0.003	0.005		0.18	0.018	0.027
		0.08	0.004	0.005		0.21	0.019	0.028
		0.09	0.004	0.005		0.25	0.021	0.030
		0.10	0.004	0.006		0.30	0.022	0.032
10,000	9,300	0.01	0.001	0.001	700	0.01	0.003	0.005
		0.02	0.001	0.002		0.03	0.006	0.008
		0.03	0.002	0.002		0.06	0.008	0.012
		0.04	0.002	0.003		0.09	0.010	0.014
		0.05	0.002	0.003		0.12	0.011	0.016
		0.06	0.002	0.003		0.15	0.012	0.018
		0.07	0.002	0.003		0.18	0.013	0.019
		0.08	0.003	0.004		0.21	0.014	0.020
		0.09	0.003	0.004		0.25	0.015	0.021
		0.10	0.003	0.004		0.30	0.015	0.023

Table 10. Adult Cohort: High Prevalence Disorders (Anxiety Disorders, Impulse-Control Disorders, General Mood Disorders), *p* = .25

Total Number (n)	Well at Time 1 (n1)	Probability (p) of Diagnosis (p1)	Standard Error (SE) (p1), deff = 0.8	SE (p1), deff = 1.7	Diagnosis at Time 1 (n2)	Probability of Remission (p2)	SE (p2), deff = 0.8	SE (p2), deff = 1.7
500	375	0.01	0.005	0.007	125	0.01	0.008	0.012
		0.02	0.006	0.009		0.03	0.014	0.020
		0.03	0.008	0.011		0.06	0.019	0.028
		0.04	0.009	0.013		0.09	0.023	0.033
		0.05	0.010	0.015		0.12	0.026	0.038
		0.06	0.011	0.016		0.15	0.029	0.042
		0.07	0.012	0.017		0.18	0.031	0.045
		0.08	0.013	0.018		0.21	0.033	0.047
		0.09	0.013	0.019		0.25	0.035	0.050
		0.10	0.014	0.020		0.30	0.037	0.053
1,000	750	0.01	0.003	0.005	250	0.01	0.006	0.008
		0.02	0.005	0.007		0.03	0.010	0.014
		0.03	0.006	0.008		0.06	0.013	0.020
		0.04	0.006	0.009		0.09	0.016	0.024
		0.05	0.007	0.010		0.12	0.018	0.027
		0.06	0.008	0.011		0.15	0.020	0.029
		0.07	0.008	0.012		0.18	0.022	0.032
		0.08	0.009	0.013		0.21	0.023	0.034
		0.09	0.009	0.014		0.25	0.024	0.036
		0.10	0.010	0.014		0.30	0.026	0.038
5,000	3,750	0.01	0.001	0.002	1,250	0.01	0.003	0.004
		0.02	0.002	0.003		0.03	0.004	0.006
		0.03	0.002	0.004		0.06	0.006	0.009
		0.04	0.003	0.004		0.09	0.007	0.011
		0.05	0.003	0.005		0.12	0.008	0.012
		0.06	0.003	0.005		0.15	0.009	0.013
		0.07	0.004	0.005		0.18	0.010	0.014
		0.08	0.004	0.006		0.21	0.010	0.015
		0.09	0.004	0.006		0.25	0.011	0.016
		0.10	0.004	0.006		0.30	0.012	0.017
10,000	7,500	0.01	0.001	0.001	2,500	0.01	0.002	0.003
		0.02	0.001	0.002		0.03	0.003	0.004
		0.03	0.002	0.003		0.06	0.004	0.006
		0.04	0.002	0.003		0.09	0.005	0.007
		0.05	0.002	0.003		0.12	0.006	0.008
		0.06	0.002	0.004		0.15	0.006	0.009
		0.07	0.003	0.004		0.18	0.007	0.010
		0.08	0.003	0.004		0.21	0.007	0.011
		0.09	0.003	0.004		0.25	0.008	0.011
		0.10	0.003	0.005		0.30	0.008	0.012

Table 11. Youth Cohort: Low Prevalence Disorders (Attention-Deficit and Disruptive Behavior Disorders, Mood Disorders), *p* = .03

Total Number (n)	Well at Time 1 (n1)	Probability (p) of Diagnosis (p1)	Standard Error (SE) (p1), deff = 0.8	SE (p1), deff = 1.7	Diagnosis at Time 1 (n2)	Probability of Remission (p2)	SE (p2), deff = 0.8	SE (p2), deff = 1.7
500	485	0.01	0.004	0.006	15	0.01	0.023	0.033
		0.02	0.006	0.008		0.03	0.039	0.057
		0.03	0.007	0.010		0.06	0.055	0.080
		0.04	0.008	0.012		0.09	0.066	0.096
		0.05	0.009	0.013		0.12	0.075	0.109
		0.06	0.010	0.014		0.15	0.082	0.120
		0.07	0.010	0.015		0.18	0.089	0.129
		0.08	0.011	0.016		0.21	0.094	0.137
		0.09	0.012	0.017		0.25	0.100	0.146
		0.10	0.012	0.018		0.30	0.106	0.154
1,000	970	0.01	0.003	0.004	30	0.01	0.016	0.024
		0.02	0.004	0.006		0.03	0.028	0.041
		0.03	0.005	0.007		0.06	0.039	0.057
		0.04	0.006	0.008		0.09	0.047	0.068
		0.05	0.006	0.009		0.12	0.053	0.077
		0.06	0.007	0.010		0.15	0.058	0.085
		0.07	0.007	0.011		0.18	0.063	0.091
		0.08	0.008	0.011		0.21	0.067	0.097
		0.09	0.008	0.012		0.25	0.071	0.103
		0.10	0.009	0.013		0.30	0.075	0.109
5,000	4,850	0.01	0.001	0.002	150	0.01	0.007	0.011
		0.02	0.002	0.003		0.03	0.012	0.018
		0.03	0.002	0.003		0.06	0.017	0.025
		0.04	0.003	0.004		0.09	0.021	0.030
		0.05	0.003	0.004		0.12	0.024	0.035
		0.06	0.003	0.004		0.15	0.026	0.038
		0.07	0.003	0.005		0.18	0.028	0.041
		0.08	0.003	0.005		0.21	0.030	0.043
		0.09	0.004	0.005		0.25	0.032	0.046
		0.10	0.004	0.006		0.30	0.033	0.049
10,000	9,700	0.01	0.001	0.001	300	0.01	0.005	0.007
		0.02	0.001	0.002		0.03	0.009	0.013
		0.03	0.002	0.002		0.06	0.012	0.018
		0.04	0.002	0.003		0.09	0.015	0.022
		0.05	0.002	0.003		0.12	0.017	0.024
		0.06	0.002	0.003		0.15	0.018	0.027
		0.07	0.002	0.003		0.18	0.020	0.029
		0.08	0.002	0.004		0.21	0.021	0.031
		0.09	0.003	0.004		0.25	0.022	0.033
		0.10	0.003	0.004		0.30	0.024	0.034

Table 12. Youth Cohort: High Prevalence Disorders (Any Mental Health Disorder), *p* = .2

Total Number (n)	Well at Time 1 (n1)	Probability (p) of Diagnosis (p1)	Standard Error (SE) (p1), deff = 0.8	SE (p1), deff = 1.7	Diagnosis at Time 1 (n2)	Probability of Remission (p2)	SE (p2), deff = 0.8	SE (p2), deff = 1.7
500	400	0.01	0.004	0.006	100	0.01	0.009	0.013
		0.02	0.006	0.009		0.03	0.015	0.022
		0.03	0.008	0.011		0.06	0.021	0.031
		0.04	0.009	0.013		0.09	0.026	0.037
		0.05	0.010	0.014		0.12	0.029	0.042
		0.06	0.011	0.015		0.15	0.032	0.047
		0.07	0.011	0.017		0.18	0.034	0.050
		0.08	0.012	0.018		0.21	0.036	0.053
		0.09	0.013	0.019		0.25	0.039	0.056
		0.10	0.013	0.020		0.30	0.041	0.060
1,000	800	0.01	0.003	0.005	200	0.01	0.006	0.009
		0.02	0.004	0.006		0.03	0.011	0.016
		0.03	0.005	0.008		0.06	0.015	0.022
		0.04	0.006	0.009		0.09	0.018	0.026
		0.05	0.007	0.010		0.12	0.021	0.030
		0.06	0.008	0.011		0.15	0.023	0.033
		0.07	0.008	0.012		0.18	0.024	0.035
		0.08	0.009	0.013		0.21	0.026	0.038
		0.09	0.009	0.013		0.25	0.027	0.040
		0.10	0.009	0.014		0.30	0.029	0.042
5,000	4,000	0.01	0.001	0.002	1,000	0.01	0.003	0.004
		0.02	0.002	0.003		0.03	0.005	0.007
		0.03	0.002	0.004		0.06	0.007	0.010
		0.04	0.003	0.004		0.09	0.008	0.012
		0.05	0.003	0.004		0.12	0.009	0.013
		0.06	0.003	0.005		0.15	0.010	0.015
		0.07	0.004	0.005		0.18	0.011	0.016
		0.08	0.004	0.006		0.21	0.012	0.017
		0.09	0.004	0.006		0.25	0.012	0.018
		0.10	0.004	0.006		0.30	0.013	0.019
10,000	8,000	0.01	0.001	0.001	2,000	0.01	0.002	0.003
		0.02	0.001	0.002		0.03	0.003	0.005
		0.03	0.002	0.002		0.06	0.005	0.007
		0.04	0.002	0.003		0.09	0.006	0.008
		0.05	0.002	0.003		0.12	0.006	0.009
		0.06	0.002	0.003		0.15	0.007	0.010
		0.07	0.003	0.004		0.18	0.008	0.011
		0.08	0.003	0.004		0.21	0.008	0.012
		0.09	0.003	0.004		0.25	0.009	0.013
		0.10	0.003	0.004		0.30	0.009	0.013

3.5 Data Collection Implications

This section describes the issues for consideration for the mental health longitudinal study related to data collection, including interviewer type and mode of administration, recruiting and training interviewers and tracers, IRB and OMB considerations, tracing and panel maintenance, and respondent incentives.

3.5.1 Interviewer Type and Mode of Administration

Interviewer Type. One key consideration for data collection is whether the instrument selected for the mental health longitudinal study is designed to be administered by clinical interviewers (CIs) or interviewers who have a bachelor's degree in psychology or related field. The qualifications needed for a non-clinical interviewer differs by the instrument to be administered, but in general, an interviewer with a bachelor's in psychology or related field would be used to administer a fully structured instrument that requires no specialized skills. For example, the CIDI is a fully structured psychiatric diagnostic interview that is designed to be used by trained interviewers who are not clinicians. However, given the nature of the questions and potential for a distressed respondent case, an interviewer with a bachelor's degree and some clinical or psychological background would be preferred for administering a structured mental health instrument. While semi-structured diagnostic interviews administered by CIs are generally considered the gold standard for making diagnoses of mental disorders, these diagnostic interviews generally are not financially or logistically feasible in large-scale epidemiological research (Jewell, Handwerk, Almquist, & Lucas, 2004).

Semi-structured diagnostic interviews, such as the SCID or K-SADS, are administered by trained clinical interviewers. Clinical interviewers require a higher level of education (MS or PhD) and rate of pay than interviewers who have a bachelor's degree in psychology or related field and need more supervision to administer and review a semi-structured (versus fully structured) diagnostic interview. Also, since symptoms and mental disorders differ for adults and children, a different set of CIs would need to be hired and trained to administer the diagnostic interview appropriate for each age group. One other consideration is that semi-structured diagnostic interviews rely on clinical judgment and therefore can be more vulnerable to interviewer biases compared to the fully structured instrument. Semi-structured instruments have been successfully administered by telephone to both adults and youth.

A fully structured diagnostic interview, such as the CIDI, can be administered by lay interviewers. As a result, the same pool of interviewers could collect data for adult and youth samples at a lower pay rate than a clinical interviewer. Structured diagnostic interviews have traditionally been administered in person, but researchers have conducted psychiatric clinical interviews over the telephone for many years as discussed in Section 3.3.7.2. If the interview is administered over the phone, a structured, centralized call center environment would be recommended with the appropriate supervision, such as on-site, masters-level supervisors with a clinical background and PhD psychologists for overseeing the supervisors and handling distressed respondent protocols.

Mode of Administration. An advantage of telephone administration is that centralized and closely supervised interviewers can carry out interviews throughout the country without the geographic restrictions or widely dispersed team required for in-person clinical assessment. The telephone interviewers do not need to be located near respondents. This centralized approach can be particularly beneficial with distressed respondents when immediate supervision is needed.

Telephone interviewing also tends to be less expensive than in-person interviewing because telephone contact does not involve travel time to and from the household. A disadvantage is that the small part of the population without telephones cannot be included when interviews are completed by telephone. Also, an increasing number of people have only cell phones. This group may be less willing to complete or finish the interview, especially a longer diagnostic interviews. However, on the 2010 MHSS, only 3% of the cases resulted in a final break-off. There is also little evidence on the relationship between interview length and response rates and breakoffs for cell phone interviews. Brick, Edwards, and Lee (2007) found that for the California Health Interview Survey (CHIS), a 30-minute telephone interview, those agreeing to be interviewed by cell phone were no more likely to break off the survey than those agreeing to be interviewed using a landline phone. It can be difficult to administer lengthy informed consent procedures and interviews over the telephone. For example, individual questions and response options cannot be as lengthy as with in-person interviews, especially since showcards cannot be used in telephone interviews. The in-person mode is more costly, but generally results in higher response rates. It is also amenable to longer interviews. Section 3.3.7.2 provides the level of agreement for two studies that compared face-to-face administration with telephone administration for diagnostic interviews.

A multimode design could also be considered, such as following up telephone administration at each wave with in-person contact for the difficult-to-contact cases. A multimode option may be more feasible with a design utilizing non-clinical interviewers because of the difficulty of locating a large number of clinical interviewers across the country within proximity to respondents. Multimode designs are often used as a way to conserve costs because less expensive modes can be used first to collect data from cooperative individuals who would respond regardless of mode. Moreover, one mode can be used to compensate for the weakness of another. For example, in-person interviewing can overcome barriers to response caused by not having a telephone number or households using call-screening devices to evade interviewers. Although multi-mode interviewing is a design option for consideration, it was not included in the cost assumptions for the design scenarios proposed in Section 4. Any design that incorporates multiple modes to collect the same data must consider the effect of the varying modes on the quality of the data collected.

3.5.2 Recruiting and Training Interviewers

Recruiting Interviewers. Over the past few years, NSDUH MHSS clinical supervisors and data collection managers have conducted extensive recruiting efforts to assemble a team of nationally dispersed CIs for the MHSS. Although the MHSS is conducted over the phone, CIs were recruited across the country to ensure availability across time zones. The lessons learned from the MHSS can be used to recruit experienced and qualified clinical interviewers for the mental health longitudinal study. If clinical interviewers are recruited for semi-structured diagnostic interviews, the selection criteria would be as follows: (1) necessary credentials,

including having a master's or doctoral degree in clinical or counseling psychology, or a medical degree with a specialty in psychiatry; (2) experience conducting clinical interviews with a background in nursing, social work, psychology, or other health-related field; and (3) child mental health experience if working on the children's mental health instrument. If lay interviewers are recruited for a structured interview, the preferred criteria would be interviewers with a bachelor's degree and some clinical or psychology background. Familiarity with the instrument would be a hiring requirement for clinical interviewers and a priority for lay interviewers.

Other key skills to consider for clinical and lay interviewers include strong conceptual skills; good attention to detail; the ability to accurately administer a complex interview protocol; the ability to develop and maintain strong rapport with respondents, including the ability to adjust interview style to competencies and the personality of the respondent; and the flexibility and capacity to work as part of a team and accept constructive feedback.

Training Interviewers. The training session for the mental health longitudinal component would be similar to the NSDUH MHSS CI training session. Prior to the session and development of materials, clinical supervisors would need to be formally trained on the selected instrument. A multidisciplinary training program built on principles of adult education would be recommended for the interviewer training, including self-study manuals, class instruction and demonstrations, and hands-on practice conducting interview mocks. Similar to the MHSS training on the SCID, a 4-day training session would be adequate for CIs to be trained on a semi-structured instrument, but an additional day of training would be recommended for non-clinical interviewers to allow time to review the questionnaire and standardized administration procedures. Both clinical and lay interviewers would be required to pass a standardized certification process.

The training session would focus on key protocols and procedures specific to contacting participants, scheduling and administering the instruments, handling distressed respondents, conducting a phone interview with adults and children, entering notes as appropriate for the diagnostic interview, and transmitting data. The training would also include instruction on basic interviewing techniques, including gaining informed consent and cooperation, probing and avoiding bias, maintaining confidentiality, and data security. General administrative tasks would be discussed as well, including instructions on the case management system and completing timesheets. Interviewers would receive one-on-one feedback and coaching from a trainer on any items missed during the certification process. The training and certification process will enable trainers to assess the interviewer's ability to follow project protocols and allow trainers to provide direct feedback and retraining prior to an interviewer beginning work, thus maximizing data quality.

Training Tracers. Another training issue for consideration is the training of staff who will carry out the tracing procedures. If new centralized or interactive tracers are needed, they would be trained on the basics of respondent tracing through the use of interactive database searches; the training would take approximately 6 hours. All tracing staff would also need to complete a 2-hour project-specific training before tracing activities begin. The project-specific training would cover topics such as sample characteristics, understanding the current contact

information and leads obtained from previous waves of data collection, confidentiality, and an overview of the tracing steps to be followed for locating sample members.

If field tracing is included in the strategies for minimizing attrition, field interviewers (or field tracers) will also receive training as needed on field tracing steps, including a brief iLearning course and teleconference with a tracing manager. If interviewers are unable to identify the respondent's current address or phone number during batch or interactive tracing, field tracers would use leads developed from previous tracing strategies. They would visit previous addresses of the sample member and revisit or re-contact solid leads from past waves. Additionally, field tracers would investigate potentially viable addresses that centralized tracers generated during the interactive tracing process. Field tracers would attempt to make contact with the respondent, provide study information, and gather phone numbers in-person. Field tracers would discuss and receive tips on field tracking strategies during training and learn how to accurately and securely record any leads identified in the field.

3.5.3 IRB and OMB Issues

Informed Consent. In order to assess the implications of IRB considerations on the proposed design scenarios, RTI project staff met with the Director of RTI's Office of Research Protection, an IRB manager, and an IRB administrator. The persons consulted felt that the longitudinal component would not need to be mentioned until the baseline interview recruitment process at the end of the main NSDUH interview, for all design scenarios presented in Section 4. This means that respondents would not be aware of the potential for inclusion in a longitudinal study when their participation is requested for the main NSDUH. This would largely mirror the current approach for the MHSS recruitment process at the end of the NSDUH interview. This design ensures that the inclusion of the longitudinal component would not affect the main NSDUH consent process and interview materials, or the response before or during the NSDUH main interview. However, IRB requirements often change over time and IRBs across different organizations vary in their requirements, and any future requirement that this information should be presented up front would lead to the inability of NSDUH to collect the longitudinal data. It is also unlikely that OMB would approve informing NSDUH participants of a second longitudinal study with multiple waves of data collection at the end of the NSDUH survey. The current MHSS study is a one-time data collection involving a very small number of respondents which is very different from the scope and magnitude of any longitudinal study.

During recruitment for the longitudinal component, respondents would need to be informed that (1) they can choose to participate in the baseline interview and refuse follow-up interviews, (2) identifiable information about the respondent would be stored for a period of time for future contacting purposes and be destroyed at the end of the study, and (3) the forthcoming interviews would be used as diagnostic tools. In addition to gaining consent from any adults selected for the mental health longitudinal study, parental or guardian consent would be needed to interview a parent about his or her child's mental health and to interview the child aged 12–17 years. After gaining parental or guardian consent, children aged 12–17 years would be asked to give their assent for participation in the interview. Children would be required to reconsent as adults when they turned 18 years old during the study.

The current MHSS includes only adults. Despite initial conversations with the IRB, the inclusion of children between 12 and 17 years of age would add complexity to the IRB approval process that may result in cost implications and other unknown implications for the NSDUH main study. For example, a distressed respondent protocol would need to be developed for children similar to the MHSS protocol developed for adults and the youth protocol developed for the NHIS SDQ Calibration Study. Also, a waiver for written parental consent for the mental health longitudinal study component would need to be requested if verbal consent was preferred. In the conduct of the NHIS SDQ Calibration Study, the National Center for Health Statistics Ethics Review Board required reporting results of the mental health interviews with participating parents; however, there was some disagreement about whether this practice was routinely necessary within mental health surveillance studies. Also, verbal parental consent and child assent by phone was approved on the NHIS SDQ Calibration Study.

Office of Management and Budget. All materials provided to respondents related to the longitudinal component would require IRB and OMB approval. OMB approval for the NSDUH is currently requested for two survey years at a time. If a longitudinal component is implemented and the number of waves is unknown at the time of OMB submission, OMB may need to provide annual approval of the longitudinal component, which would have additional cost implications for the NSDUH.

3.5.4 Tracing and Panel Maintenance

The need to trace sample members in subsequent waves of a longitudinal study can be triggered by a move or change in telephone number by the household or selected individual. For example, 12.5 percent of the U.S. population moved in 2010, according to the 2010 Current Population Survey. For data quality purposes, it is important to minimize attrition by identifying and locating sample members who have moved or changed phone numbers because those who move are likely to be different from nonmovers. Attrition can also impact response rates for follow-up waves. Location propensity can be affected by many survey design factors, including study population, tracing rules or steps employed, time and effort expended in tracing movers, time between waves, contact information collected on sample members, and contact maintained with sample members between waves (Couper & Ofstedal, 2009).

Minimizing panel attrition at each follow-up wave will require a robust tracing strategy. Recent literature indicates that obtaining high locate rates in a longitudinal study is possible when a tailored locating approach involving multiple searches of databases, telephone calls, and in-person visits is used (Haggerty et al., 2007). Add Health Waves III and IV used a combination of tracing techniques, including batch tracing, respondent mailing before data collection, interactive (or centralized) tracing, and field tracing. The Wave IV tracing plan was further enhanced by including additional batch tracing services, respondent address updates being mailed after batch tracing, multiple address update options for respondents, additional pre-data-collection interactive tracing, and additional batch tracing during data collection, resulting in a 92.5 percent locate rate for Wave IV in comparison to an 87 percent locate rate in Wave III. Table 13 provides a comparison of the percentage of successful locates for Wave III and IV by each tracing step (Meehan et al., 2009).

Table 13. Add Health Wave III and Wave IV Comparison—Percentage Located, by Tracing Step

Tracing Step at Which Sample Member Located	Wave III	Wave IV
Advance tracing (batch tracing or interactive tracing prior to data collection start)	37%	66%
Interactive tracing (after start of data collection)	26%	10%
Located in the field	12%	14%
Located via other tracing step	0%	3%
Located noninterviews, unknown source[a]	12%	0%
Not located	13%	8%

[a] In Wave III, located noninterviews were not accounted for in the locate percentages; thus, in 12 percent of Wave III cases, the tracing step cannot be identified that led to locating the respondents of located noninterview cases.

Creating a strong base of tracing information at baseline will be essential to minimizing attrition in follow-up waves. Similar to the MHSS, only the minimal contact information (i.e., first name and phone number) would be collected at the end of the main study interview to contact the respondent for the baseline interview, which would occur up to 4 weeks later. In the unlikely event that OMB and any IRB permit not informing the respondents ahead of time, the informed consent statements for the main NSDUH interview would not change, because no more information would be collected at the end of the main study interview than current procedures for the MHSS. At the end of the baseline interview for the mental health longitudinal study, it would be useful to collect detailed contact information from respondents—including their home address, email address, telephone numbers, and dates of birth—to locate them for future waves. It is also helpful to collect contact information for friends or relatives who will know how to reach the respondent at a later date. Friends and relatives are often the best source for finding respondents who have moved or changed phone numbers in follow-up waves. The contact information would be keyed into a separate system or module than the interview data by tracing staff and interviewers, and kept separate from the interviewing data using a different identification number. This would enable the link between the interview data and contacting information to be broken at any time. This would be a very large departure from current NSDUH procedures where such detailed information are not collected from respondents. Systems would need to be modified to handle these data and would modify the current privacy impact assessment that is currently done. A system of records would need to be established for the collection of such personally identifiable information. These would lead to additional administration costs both for the government and contractor.

Social Security numbers (SSNs) are the best guarantee of a reliable match when searching commercially available databases. When an SSN was available for tracing on Add Health Wave IV, a significantly higher percentage of Wave IV cases were located (Meehan et al., 2009). The sensitivity of the data collected would impact the cost considerations as well as OMB and IRB approval decisions. For example, the collection and storage of SSNs may need to be protected by Federal Information Processing Standards (FIPS) moderate-level security controls, which will increase associated data management costs. Since collecting respondents' social security numbers (SSN) raises privacy and data security concerns, alternatively, the last four digits of a respondent's social security number could be collected, which will be less sensitive but will still assist in tracing efforts. However, it is unknown whether approval to collect the entire or partial social security number would be granted to SAMHSA.

Table 14 provides a summary of tracing strategies for consideration to minimize attrition in follow-up waves.

Table 14. Tracing Strategies for Minimizing Attrition

Tracing Method	Description
Acquisition of Contact Information at Baseline	• Collect contact information including address, phone numbers, e-mail address, and SSN during the baseline interview • Collect information on contacts who would always know how to reach the sample member if he or she were to move • Provide solid information to be used in subsequent tracking strategies at follow-up waves
Batch Tracing	• Match entire groups of cases quickly at a relatively low cost through electronic database searches (e.g., TransUnion, Accurint, Telematch) (This is the most cost-effective way of locating the most recent contact information for a large group of participants.)
Panel Maintenance Mailing	• After completing batch tracing, send letters to all participants inviting them to update or confirm their contact information; consider providing a small incentive to increase compliance • Provide multiple methods for updating contact information, including a locator update form, toll-free number, e-mail address, and website • Include in the panel maintenance mailings a study magnet that sample members can retain to easily locate the study toll-free number and other study contact information • Target returns from the post office without a forwarding address for additional tracing steps (Contact information from post office returns with a forwarding address would be updated.)
Interactive (or Centralized) Tracking	• Refer cases not located through the efforts above to well-trained staff specializing in tracing and locating sample members of all types • Send cases for individual assessment and review by tracers (Tracers may use U.S. credit bureaus [Experian, Equifax, and TransUnion] as well as consumer information databases and public records.)
Social Networking Websites	• Utilize social networking websites to reach younger, more mobile populations
Field Tracing	• Have field interviewers use address leads developed through the methods described above • Reserve field tracing for the hardest to locate cases because of its greater cost

3.5.5 Respondent Incentives

Incentives would serve an important role in maintaining acceptable response rates across multiple survey requests in a NSDUH mental health longitudinal study. Maintenance of these response rates serves to maintain sample size, potentially limit nonresponse bias on survey estimates, and reduce field work costs. To address the burden that multiple survey requests impose on respondents, the incentive amount for the baseline survey should be appropriate for the length and burden of the interview. For example, if the length of the baseline panel survey for an adult averages 60 minutes, a $30 incentive for the baseline interview would be an appropriate token of appreciation. This would be consistent with the current incentive provided on the main

NSDUH survey, which is the most likely frame of reference for respondents evaluating the request to participate in the mental health longitudinal survey. Assuming that the 6-month follow-up assessment is about 30 minutes, a $15 incentive would be appropriate given that it should take about half the time required for the baseline interview.

If the adult baseline interview requires additional modules, such as the addition of a lifetime component and the full psychotic disorders assessment to the current MHSS SCID assessment, and takes approximately 90 minutes on average, an incentive of $45-50 would be appropriate. The NCS-R paid respondents a $50 incentive. It required a minimum of 90 minutes to complete among respondents who reported no lifetime disorders, an average of approximately 2 hours and 30 minutes among people with a history of disorder, and as long as 5 to 6 hours among respondents with a very complex history of many different disorders (Kessler et al., 2004). In summary, when considering the incentive for the baseline interview for the mental health longitudinal study, it will be important to consider respondent burden or the length of the follow-up instrument in comparison to the length of the main interview and incentive provided to main study respondents.

For the NCS-A sample, each respondent (one adolescent and one parent or parent surrogate in each household) was given a $50 incentive for participation (Kessler et al., 2009). The adolescent interview in the NCS-A was quite long, with an average length of 2½ hours, and the parent component was approximately 45 minutes. Given that the proposed design requires a parent and child interview pair for a complete interview, an incentive for the child and parent should also be considered for the NSDUH mental health longitudinal component.

Singer and Bossarte (2006) state that designs requiring continued respondent participation have an inherent justification for offering higher incentives. This justification should carry over across the waves of the longitudinal design. Research shows that the study design should entertain increasing the incentive amount at each wave of the study. In their review of the use of incentives in panel surveys, Laurie and Lynn (2009) report that surveys typically increase incentives over time to keep them in line with the cost of living. Their own analysis led them to conclude that "even a small increase in the value of an incentive on a mature panel can bring a significant improvement in response rates" (Laurie & Lynn, 2009, p. 230). More recently, Rodgers (2011) reports on an experiment with the Health and Retirement Study (HRS), in which respondents would receive the same incentive as before ($20) or an increased incentive (either $30 or $50). Those offered the $50 incentive had higher response rates in subsequent waves than those who continued to be offered the $20 incentive. Furthermore, increases in incentives were also associated with improved data quality in terms of lower item nonresponse and less of a tendency to give answers that would shorten the survey. Following this reasoning, it is recommended that the NSDUH mental health longitudinal study should offer an increased incentive of at least $10 with each annual follow-up. With the exception of the baseline study, each of these incentives will be promised rather than prepaid as with the MHSS. Respondents will receive the incentive upon completion of the interview. This model of using promised incentives would mirror that used by the Panel Study of Income Dynamics, a longitudinal telephone study. The incentive could be sent via a check addressed to the respondent or a gift card with an activation code provided only to the respondent.

There is a general consensus that prepaid incentives are more effective at increasing response rates than promised incentives for the initial attempt to complete a survey (Cantor, O'Hare, & O'Connor, 2007; Gelman, Stevens & Chan, 2003; Singer, 2002). However, there is comparatively little information on whether there are differences in response rates between prepaid and promised incentives in a panel study and even less information on whether or not a change in the method of providing an incentive from one wave to the next has any effect on response rates. Castiglioni, Pforr & Krieger (2011) report on an experiment in which respondents in a pilot study for the German Family Panel were randomly assigned to three incentive conditions: prepaid, promised and no incentive. At each wave, the incentive was provided in the same manner within each treatment group. The survey became more burdensome with each subsequent wave so across all three groups, response rates declined. At the initial wave, response rates were high and there were no statistically significant differences in response rates between any of the three conditions. At the second wave, the promised incentive condition had a higher response rate than either the prepaid or no incentive groups. At the third wave, the response rates for the prepaid and promised incentive groups were both higher than those in the no incentive group. Across all three waves, it was found that the promised incentive group had a higher response rate than the no incentive group and the prepaid incentive group had a response rate that was not statistically distinguishable from the rates for other two groups.

3.6 Data Quality

The study design and instrumentation tasks of a NSDUH mental health longitudinal study will strive to reduce the influence of measurement error and thus improve data quality. Data quality can be improved by using appropriate reference periods and introducing questionnaire design features that will aid respondents in providing retrospective reports of mental health statuses and events. Much of the literature about recall in surveys applies to longitudinal surveys in the same way as it applies to cross-sectional surveys. Respondents to longitudinal surveys engage in the same response processes as respondents to cross-sectional studies. However, follow-up questionnaires have the advantage of using data collected in previous waves to remind respondents about previous responses and events. For instance, stable demographic information can be prefilled from earlier rounds of data collection and respondents can confirm the information in later rounds. Similarly, information collected for tracing purposes can be confirmed with respondents. These basic fills were assumed in the cost estimate provided for the various design scenarios in Section 4. The questionnaire development process can evaluate the extent to which other preloads could anchor reference periods or reduce respondent burden. The questionnaire design process should entertain employing dependent interviewing, which uses data collected at previous waves to word questions or route the respondent through the interview. (Uhrig and Sala, 2011)

3.6.1 Reference Periods

An appropriate reference period will aid respondent recall while reporting retrospective events. A respondent's ability to retrieve the information necessary to accurately respond to a question is adversely affected by longer reference periods. The longitudinal study should collect retrospective reports of mental health status in a reasonable time frame that will allow respondents to avoid the normal forgetting process, recall the dates of these events accurately, and minimize the intrusion caused by multiple events in the same period. Lengthy recall periods

will potentially increase the number of "Don't Know" responses, as well as incur larger numbers of response errors.

Pierret (2001) examined the impact of the National Longitudinal Study of Youth and the Panel Study for Income Dynamics moving from a 1-year reference period to a 2-year reference period. He found that, when asked about events that occurred in the past 2 years, respondents often neglected to report events that were atypical for them. Short periods of unemployment, brief receipt of benefits, and brief jobs were underreported.

When assessing the feasibility of changing the National Crime Victimization Survey's (NCVS) recall period from 6 months to 1 year, Lee and Carr (2009) introduce a number of methods for improving respondent recall for a longer period. One such method is using enhanced contextual priming, which are items that prime the respondent to think about various characteristics of crimes. For instance, the questionnaire contains questions about feeling safe inside and outside the home, while traveling, and while with strangers and acquaintances. The inclusion of these questions provides the respondent time to think about the topic and context for the upcoming questions. The authors note that recall of crime victimization often functions differently than recall of employment or program participation because it is typically a rare event.

3.6.2 Bounding

To aid respondent recall of events, the diagnostic interviews should provide temporal boundaries for recall. The assessment timeframe should be included in all instrumentation (e.g., "In the past 12 months, that is since September 12, 2010…") Providing these boundaries can reduce "forward telescoping," which occurs when respondents include events that occurred before the reference period in reports of what happened during that reference period (Tourangeau, Rips, & Rasinski, 2000). Forward telescoping results in an overestimation of events that occurred during the reference period. Neter and Waksberg (1964) found that respondents reported a larger number of events occurring in reference periods that were not bounded, compared to those that were. In other words, when the questionnaire reminded respondents of events that they had reported in prior rounds of interviews, respondents were better able to restrict their current reporting to the appropriate time period.

A similar practice has been employed in the administration of the National Crime Victimization Survey (Lee & Carr, 2009). Previous study designs called for data collected in the first wave of the study to go unanalyzed and be used only for bounding. Subsequently, these data were analyzed and statistical corrections were made to correct for bias introduced by telescoping. Unfortunately, unlike in the main study, the longitudinal interviews could not continue to make use of the reference date calendars which have been shown to aid recall and underscore the specific dates that make up the reference period since the longitudinal interviews will be conducted over the phone.

3.6.3 Panel Conditioning

One data quality concern that is unique to panel surveys is that the act of asking questions may itself lead to changes in behavior, attitudes, or emotional states that are then reported in

subsequent surveys. Panel conditioning or "time in sample bias" is not a well understood phenomena for the mental health measures that may be used on the NSDUH longitudinal mental health survey. French and Sutton (2010) provide a review of findings on the related concept of measurement reactivity in which the act of measurement induces changes in the subjects being measured. Their review covers effects on health behaviors, emotions (including depression and anxiety), and cognition. With respect to bias from measurement reactivity, they conclude that "Given the current state of the evidence, it is probably wise to regard such bias as a pervasive problem, especially when assessing emotions…" (French & Sutton, 2010, p. 464).

Wilson and Howell (2005), using data from the Health and Retirement Study (HRS), report that respondents in an initial wave who were asked if a doctor ever told them they have arthritis reported higher levels of arthritis 2 years later than the general population. Wilson and Howell (2005) speculate that one reason for panel conditioning effects in diseases and other health conditions is that participation in a survey on health items may lead survey participants to make inquiries about the conditions asked about in the survey or mention symptoms they may not otherwise mention. They also raise the possibility that over the long term, participation in a health survey may increase the respondent comfort in the process, which may in turn affect self-reporting of health conditions (Wilson & Howell, 2007).

Few studies appear to have examined the potential for panel conditioning effects for measures of depression and psychological distress. Perhaps the most relevant findings are reported by Warren and Halpern-Manners (2011) in their review paper on panel conditioning in sociology. They note that one option to examine the potential for panel conditioning effects is to drop key items of interest for a random subsample of respondents in the initial baseline wave and then ask these key items for all respondents in the subsequent wave. This approach to studying panel conditioning has the advantage of ruling out panel attrition as a potentially confounding cause of differences in responses over time. An unintentional version of this approach was used in the Wisconsin Longitudinal Survey, in which an initial telephone survey was followed up with a mail survey a few days or weeks later. For the initial telephone survey, respondents were asked questions about lifetime depression except for a subsample of respondents who were not asked these questions. In the follow up mail survey, all respondents were asked 20 items from the Center for Epidemiologic Studies Depression Scale (CES-D). For three of the 20 items on the follow-up survey, there were statistically significant differences between responses by those who were asked depression items on the initial survey and those who were not. Those that did not respond to the phone survey items about depression were more likely to report being happy, enjoying life and were less likely to not be able to "get going".

Panel conditioning effects for items on depression may be limited to follow-up surveys that are conducted soon after the initial survey. The time lag in the Warren and Halpern-Manners (2011) study varied from a few days to a few weeks. Longwell and Truax (2005) report on an experiment in which a sample of university participants was administered a web survey that included items from the Beck Depression Inventory (BDI-II). Respondents were randomly assigned to weekly, monthly, and bimonthly administrations of the survey for a 9-week period. Those who were interviewed on a weekly basis showed a pattern of declining BDI-II scores while those interviewed monthly or bimonthly showed no such changes over time.

3.6.4 Seam Effects

Seam effects occur in a number of longitudinal surveys, and they are particularly present when data are collected every few months but respondents are asked to report on each month separately since the prior interview. Seam effects bias measures of change in longitudinal surveys. When seam effects occur, the level of change in the data are greater when reports for two time periods come from different interviews. (Jäckle 2008) Data that suffer from seam effects will typically show consistent patterns across each month within a single reference period, but changes in consecutive months across interviews. Rips, Conrad, and Fricker (2003) propose a number of causes of seam effects. The first is that respondents are less able to remember months earlier in the reference period clearly. When respondents are asked to report about the last 4 months, the accuracy of their reports for the most recent month will be higher than reports of the most distant month. This differential accuracy creates an erroneous pattern in data across the months. Next, respondents may engage in "constant wave responding," where they report constant behavior in all months of the reference period in order to simplify their reporting task. When reports differ across reference periods, this causes a change at the "seam" of the waves.

While the specific questions of the NSDUH mental health longitudinal study have yet to be determined, seam effects appear unlikely to pose a threat to data quality. Research on seam effects has been based on questions in which respondents are asked to provide information for specific time intervals within an overall period of recall. The NSDUH mental health longitudinal study would not ask questions in a way designed to determine the presence or absence of disorders for specific time intervals (e.g. on a monthly or quarterly basis.)

If questions are included in which respondents are asked to recall by period, this report has already noted some techniques for reducing the effect. In an attempt to correct for seam effects due to difficulty of recall and constant wave reporting, Rips, Conrad, and Fricker (2003) experimented with a number of remedies. Dependent and bounded interviews reduced the size of the seam effect but did not rectify errors in reporting. A recommendation was also made for the use of event history calendars as a tool to combat this source of reporting error.

3.6.5 Coverage and Response Rate Issues

The proposed longitudinal study does not include those who did not complete the NSDUH in English (i.e., it excludes those who completed the NSDUH instrument in Spanish). About four percent of the NSDUH sample completes a Spanish questionnaire each year. Exclusion of this population may affect national estimates, and particularly, estimates for Hispanics. The other issue is accounting for nonresponse on the NSDUH—the baseline survey on the longitudinal study will have to account for screener and interviewer nonresponse in addition to nonresponse to the longitudinal study. Further discussion of this topic is included in Section 4.3.

3.7 Trend Analysis

There are two general levels of trends of interest in studies examining prevalence, incidence, and course of mental disorders: trends in the population and trends within individuals (course of disorders).

Population-level trends measure changes in prevalence or incidence at a population level, usually in periodic intervals. For example, NSDUH measures changes in drug use in the population annually. The most direct method of assessing population-level changes uses a repeated cross-sectional study design, which consists of a single assessment that includes a new, nationally representative group of people surveyed periodically. However, although repeated cross-sectional studies are good for obtaining population-level trends over time, they are not appropriate for following individuals over time to examine developmental changes or for assessing potential causal factors. (Because both exposure and outcome are measured concurrently in a cross-sectional study, it is difficult to discern temporal sequence of cause and effect.) Conversely, longitudinal designs can assess changes within people over time and are better suited to answer questions of individual-level course and causal factors, but do not directly measure population-level trend. This is a trade-off between the two study types.

Any measure of trend (population or individual level) will be impacted by changes in the criteria being used to measure that trend. Within this study design there are two general points of measurement interruption: changes in assessments between youths and adults, and changes in disorder criteria as measures change to reflect criteria put forth in the new DSM-V (currently estimated for publication in May of 2013).

The change in diagnostic measurement between youths and adults has already been discussed, and diagnostic criteria changes from DSM-IV to DSM-V mid-study would create similar difficulties. In this case a "dual coding study" is one possible solution but would add substantial response burden to participants. Moreover, although this approach may identify the sources of discontinuity, it does not eliminate the discontinuity. Alternatively, the study could keep the existing measurement instruments and stipulate that findings are specific to DSM-IV diagnostic criteria, as was common during the transition from DSM-III-R to IV. It should be noted that revision and validation of clinical measures for the new DSM-V criteria may take some time after final publication of the DSM revision. Therefore it cannot be determined when new versions of the SCID or any of the other diagnostic measures would be available.

3.8 Effect on Main Study

The effect of a mental health longitudinal component on the NSDUH main study will depend on the final design of the study. To minimize the impact on the NSDUH main study, it would be ideal if the longitudinal sample could be drawn from the NSDUH main interview and the longitudinal study could be introduced at the end of the NSDUH interview to selected participants. An introduction of the longitudinal component prior to the end of the main interview in any manner would most likely impact main study response rates or estimates. Since OMB and quite possibly any IRB, is not likely to approve waiting until the end of the NSDUH to inform respondents about the mental health longitudinal component, the longitudinal mental health study is likely to have a strong impact on the current NSDUH. All contact materials will

have to be modified to describe this additional survey. Respondents are more likely to refuse to participate given the possibility of a much higher burden than the current NSDUH which will affect response rates and impact the bias to NSDUH estimates. Field staff will have to work harder to convince respondents to participate thus increasing field staff burden, which in turn may make it harder to recruit and retain NSDUH main study field staff.

As discussed before, the addition of a mental health screening question to the NSDUH screening instrument would lengthen the screening time and potentially appear intrusive to respondents; this could impact participation in the main NSDUH interview. One discussed design includes the addition of the SDQ or other mental health measure to the main NSDUH questionnaire to oversample youths at risk for mental health problems, similar to utilizing the K6/WHODAS scores to select adult respondents for the mental health longitudinal study. The 25-item version of the SDQ takes an average of 5 minutes. It is estimated that the 6-item version would add 1 to 2 minutes. The 6-item SDQ is administered to the parent in the NHIS, but the 25-item SDQ can be administered to a parent or child. However, literature suggests that SDQs completed by parents (and teachers) are generally better predictors (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000), leading to another issue for consideration: whether to have 12 to 17 year olds complete the SDQ themselves during the audio computer-assisted self-interview (ACASI) portion of the questionnaire or to have a parent, guardian, or other primary caregiver complete it when serving as a proxy during the back CAPI. Of 12- to 17-year-old respondents who completed the 2010 NSDUH, 10.71% answered the proxy questions themselves. Of those who had proxies, over 83% were a parent. If a parent, guardian, or other primary caregiver was required to be present in the household to complete the SDQ portion of the interview, this requirement would strongly negatively impact main study response rates for youth when it is difficult to schedule an interview when all parties were available. There could also be context effects to the main questionnaire based on where the SDQ is inserted in the questionnaire which could lead to a break in trends for the affected items on the NSDUH main interview. Similarly, there could be context effects on the SDQ items because of the NSDUH questions asked beforehand. Thus, if the addition of the SDQ to the main NSDUH questionnaire were considered, the consequences would need to be further researched and tested. Costs associated with administering and programming the SDQ items within the main questionnaire were included in the cost assumptions presented in Section 4, but not further researching and testing the SDQ. The use of current NSDUH youth items such as mental health treatment and MDE in the longitudinal study sample design would not have any of these adverse effects.

If the mental health longitudinal component were implemented in NSDUH, a sampling issue to consider is that field interviewers return to new selected dwelling units in same segment as the previous year to conduct main study interviews under the current design. This means that main study interviews could be conducted within the same quarter in the same segment that mental health longitudinal study respondents are being contacted for the annual follow-up.

For data management, no one of the impacts on the main study would be related to the data security issues related to the collection of future contact information at the end of the main NSDUH interview. This contact information would be stored separately from the main questionnaire data, but would require movement to FIPS level moderate security controls. For the longitudinal component, systems would need to be developed to interact with tracing

components, confirm and store respondent locator information, and manage the longitudinal cases.

The NSDUH editing and coding team does not currently edit the SCIDs for the MHSS. This process is completed by experienced clinical supervisors. The editing procedures for the mental health longitudinal component would be dependent on the selected instrument, but would likely be similar to the MHSS. If the SDQ were to be added to the NSDUH, then one effect on the main study regarding editing would be that those questions would need to be edited. No impact is anticipated to the main study for weighting. New weights would be required for the longitudinal study which would be developed after the creation of the analysis weights on the NSDUH.

From the government's perspective, the longitudinal study would significantly increase the scope of work and the staff needed to manage such a complicated project. There would also be additional implications regarding how the NSDUH is contracted, when and how OMB clearance is requested, etc. This report does not cover these topics; however they are not trivial and must be considered prior to making any decisions.

For the national findings report, assuming the current format remains, the mental health longitudinal study would likely have minimal effect on the national findings report because the national findings report focuses on substance abuse and related topics. For the mental health findings report, the longitudinal design could affect the report by including discussion of methodological issues about the study in an appendix as is currently done for the MHSS; or the longitudinal component could affect the report in a broader way if the decision were made to include a chapter and reports on results from the longitudinal mental health study. The decision could also be made to develop a separate national report on the longitudinal component similar to what is done for the young adult sample in the Monitoring the Future (MTF) study, which is a follow-up sample from the high school senior sample. This would not have any impact on the main study analytic reports, other than potentially providing additional topics for reports.

Depending on the final design, one issue for consideration would be the frequency of reports. Results could be reported annually, or more frequently with the 6-month assessments. The cost estimates for the design scenarios presented in Section 4 assume that reports and detailed tables would be provided annually and at the end of the longitudinal study. The costs for reports were not included in the proposed budget.

3.9 Dissemination

Because CBHSQ is a federal statistical unit, the requirements of the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002 must be followed. This act establishes and requires the use of confidentiality protections for information collected for statistical purposes. CIPSEA also describes the means by which data collected by a federal statistical agency may be shared with other federal agencies.

Mental health data collected via NSDUH must be handled according to CIPSEA requirements. Consequently, because data collected from the main NSDUH survey are used for statistical purposes only, data collected for the mental health longitudinal study must be used

only for statistical purposes. Furthermore, data collected for the mental health longitudinal study must be reviewed for disclosure risk and statistical methods employed to reduce the likelihood of disclosure or respondent identity or attributes.

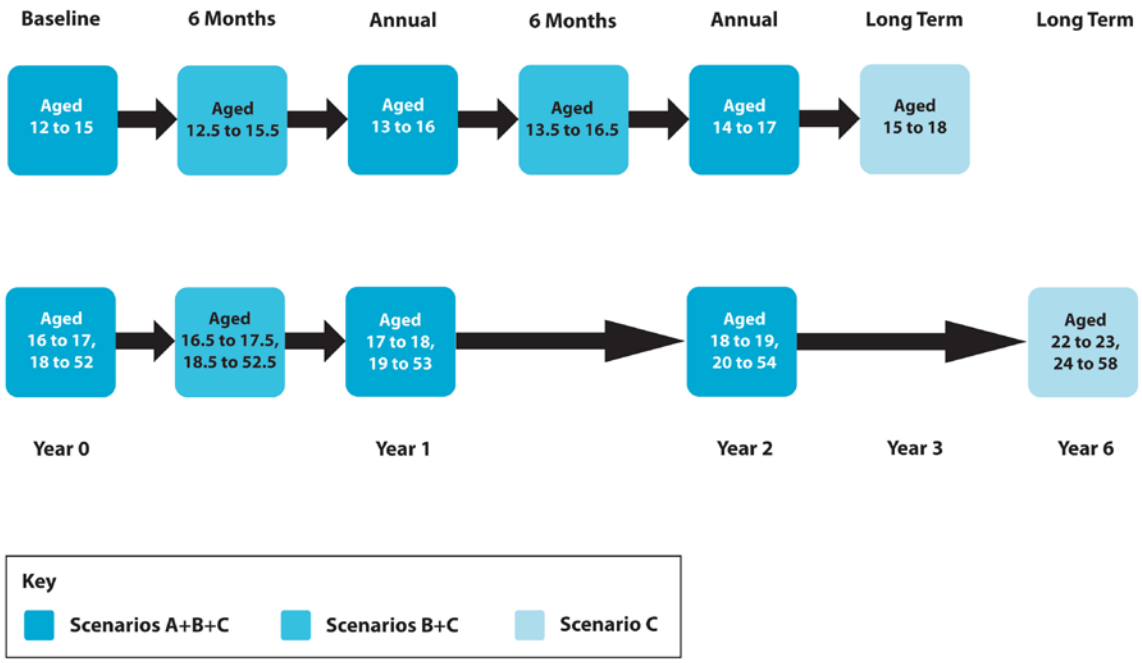
Because there will be many fewer participants in the mental health longitudinal study than in NSDUH, the nature of linked longitudinal files, and because the prevalence of various mental health conditions is expected to be small, it is unlikely that data collected for the mental health longitudinal study could be treated for disclosure risk and yet retain a high degree of data utility. Consequently, data collected for the mental health longitudinal study would not be available through public-use files and instead could be made available to end users via a set of detailed tables. The contents of these detailed tables must be reviewed for disclosure risk, and, where necessary, data values will be suppressed or otherwise perturbed to reduce the risk of disclosure of respondent identity or attributes. This process was included in the cost estimate presented for the design scenarios in Section 4.

Because these data are collected under CIPSEA, there will be restrictions regarding NIMH access to these data. NIMH will not be able to make these data available to any researcher. Under the law, that process will have to be "controlled and supervised" by CBHSQ. The process of disseminating restricted-use data is intensive in cost and use of staff time. It requires the presence of a restricted-use data dissemination infrastructure and activities such as the submission of analysis and data security plans, conducting site inspections and disclosure review, etc. These costs are not included in the cost estimates presented in Section 4.

4. Longitudinal Design Scenarios

To develop the mental health longitudinal design scenarios, RTI weighed the analytic goals of the study against the feasibility of conducting such a study and the impact of the design on the main study and costs. Literature on estimated prevalence and incidence rates and life course of mental disorders was also researched to inform the number and length of waves and sample size. Based on these considerations and research, three longitudinal design scenarios were developed to meet the analytic goals of the study. Figure 1 provides a summary graphic of the three longitudinal design scenarios. All design scenarios assume that a sequential cohort design will be used as described in Section 3.2.2, and the adult and youth sample for the longitudinal component will be selected based on their respective K6/WHODAS and SDQ scores in the NSDUH main interview as described in Section 3.4.1.

Figure 1. Longitudinal Design Scenarios



4.1 Design Scenarios

Appendix A provides separate diagrams of the three longitudinal design scenarios shown above in Figure 1. Each scenario builds off the previous scenario by providing additional assessments of life course. All three scenarios would provide estimated prevalence and incidence rates by collecting data at the baseline interview plus two annual assessments (Scenario A). Scenario A also provides the opportunity to gather some course information across three data points. However, the sample sizes must be large enough to detect course information. Because only a small proportion of the sample will have a diagnosed disorder at baseline, the sample size becomes very small when limiting it to those with the diagnosis of interest. For this reason, it is assumed that Scenario A will have a smaller sample size than Scenarios B and C and will provide only estimated prevalence and incidence rates.

In addition to the annual follow-ups, Scenario B includes one or two brief 6-month assessments to gather data on mental disorders with short-term course, such as major depression or episodic disorders. The 6-month assessments would also be useful for gathering data on transient or sporadic factors that might be relevant to understanding functional impairment, as well as risk or protective factors. In addition, the 6-month assessments will help to reduce recall bias by shortening the recall period. As discussed in Section 3.2.3.3, findings from studies using annual follow-up waves have revealed substantial change from year to year during adolescence, as well as the emergence of concurrent comorbidity of two or more disorders in later adolescence (Angold, Costello, & Erkanli, 1999; Costello, Foley, & Angold, 2006; Rowe et al., 2010). This suggests rapid fluctuation among children with psychiatric diagnoses. To address this, two 6-month follow-ups are proposed among the subsample of youths aged 12 to 15 years old that exhibited a mental disorder at a previous assessment.

Finally, Scenario C includes a long-term follow-up after baseline of 3 years for respondents aged 12 to 15 and 6 years for respondents aged 16 to 52. This final scenario targets disorders with longer term courses. Some disorders have chronicity as a feature of the diagnosis (e.g., dysthymia), and others have data suggesting routinely long periods of illness (e.g., agoraphobia, eating disorders). This final scenario would also gather data on remission and recurrence.

Section 3.2.3 provides additional information on each assessment type. The design scenarios propose to combine adolescents aged 16 or 17 with the adult follow-up pattern. This decision was based on literature demonstrating a stabilization of many adolescent frequent disorders after the age of 16 (Costello et al., 2003).

4.2 Instrument Options

Conducting a longitudinal mental health study will require procuring, developing, or identifying measures designed to target the analytic goals of the study. Optimally, the instruments for the youth and adult samples would be very similar to provide assessment continuity and comparability of the results across time. When forming recommendations for instrumentation, the domains of interest and compatibility of measures used to assess youths and adults was considered along with the psychometric properties and administration time, format (i.e., structured or unstructured), and mode. Instruments considered for a 6-month assessment would also need to be comparable to the baseline and annual follow-up interviews.

Table 15 provides two diagnostic interview instrument options (for use in the clinical follow-ups) for consideration: (A) Composite International Diagnostic Interview (CIDI) for respondents aged 12 or older and (B) the Structured Clinical Interview for DSM-IV-TR Axis I Disorders Non-patient Edition (SCID-I/NP) for respondents aged 18 or older and the Kiddie-Schedule of Affective Disorders and Schizophrenia (K-SADS) for respondents aged 12 to 17. After considering similar studies previously conducted and costs, it is assumed that both options would be administered over the phone. Abbreviated versions of these instruments would be used for the 6-month follow-up assessments. It is also assumed that parent and child versions of the instruments would be completed for 12 to 17 year olds. Parents tend to be more accurate reporters of youth externalizing behavior problems (e.g., conduct disorder, attention-deficit/hyperactivity disorder), and adolescents tend to be more accurate reporters of

internalizing problems (e.g., anxiety or depression) (Jensen et al., 1999). Consequently, the research literature emphasizes the importance of conducting both parent and adolescent interviews to generate accurate estimates of youth mental disorders. Section 3.3.7 provides additional details on the candidate instruments.

Table 15. Advantages and Disadvantages of Instrument Options

Option	Instrument(s)	Advantages	Disadvantages
A	12 or older: CIDI	• well-researched; very good adult psychometric properties • administered by lay interviewers, thus less expensive than semi-structured interviews • same pool of interviewers could collect data for adult and youth samples • inclusion of essential information not assessed by the SCID, including disorders often identified in childhood (e.g., attention-deficit/hyperactivity disorder, oppositional-defiant disorder, conduct disorder, separation anxiety disorder) and disorders diagnosed among adults (e.g., personality disorders, premenstrual disorder, and pathological gambling) • assessment of physical comorbidity, treatment services, pharmacoepidemiology, and the risk factors of social networks, childhood experiences, family burden, and childhood demographics • more seamless evaluations and better comparability of results during transition from adolescence to adulthood	• procedures, protocols, and training materials would need to be adapted from the MHSS for the CIDI • lengthy interview; modules could be dropped to reduce respondent burden • traditionally administered face-to-face; maintaining respondent's attention over the phone might be difficult due to lengthy administration time • no assessment of all of the low prevalence diagnoses for adults (However it does screen for general psychotic symptoms.) • previously administered to youths aged 13 years old or older, but not those 12 years old • psychometric properties of the adolescent version of the CIDI are not yet well established
B	18 or older: SCID; 12–17 years: K-SADS	• established administration procedures, protocols, and training materials for the SCID from the MHSS; could also be adapted for the K-SADS • consistency in assessments of adult disorders and estimates between the MHSS and the longitudinal study • assessment of low-prevalence disorders of adjustment disorder and all of the major psychotic disorders • shorter administration time in comparison to the CIDI • with K-SADS, more evidence of reliability and validity compared with the CIDI • successfully administered by telephone to both adults and youth on previous studies	• limited domain assessments in comparison to CIDI • higher rate of pay and more supervision required for clinical interviewers • different set of clinical interviewers would be needed for the SCID and K-SADS • reliance on clinical judgment; can be more vulnerable to interviewer biases • two different instruments might lead to difficulties comparing the transition from adolescence to adulthood

CIDI = Composite International Diagnostic Interview; K-SADS = Kiddie-Schedule of Affective Disorders and Schizophrenia; MHSS = Mental Health Surveillance Study; SCID = Structured Clinical Interview for the DSM-IV.

4.3 Sample Sizes

Tables 16 and 17 provide the starting sample sizes and expected response rates (that account for both attrition and nonresponse) at each phase for adults and youth, respectively. The response rates for the follow-ups are among the baseline sample. The 6-month follow-ups are completed only for persons with the diagnosis at baseline and the first annual follow-up (if applicable). A nonrespondent from one follow-up will still be contacted and interviewed at the next follow-up. Thus, the response rates for the longitudinal study are not cumulative. The overall response rate is computed as the product of the main study screening and interview response rate and the final longitudinal response rate (i.e., the number of final respondents divided by the number of persons selected for longitudinal follow-up). The response rate assumptions were obtained from the 2010 NSDUH and MHSS and supplemented with results from longitudinal literature and experience administering longitudinal telephone interviews. The measures used to compute the desired final sample sizes are specified in Tables 8 through 12 in Section 3.4. The youth cohort response rates assume that a complete parent and child pair interview is required for an interview to be considered complete.

When accounting for the expected attrition and nonresponse across both NSDUH and the longitudinal study, the overall response rate is very low. The current NSDUH has approximate weighted screener and interview response rates of 89 and 74 percent respectively. Even with an expected baseline interview response rate of 67 percent (based on experiences on the MHSS and other longitudinal studies), the starting response rate on the NSDUH longitudinal study will be approximately 44 percent. Therefore nonresponse weighting adjustments will have to account for approximately 70 percent of the nonresponding cases. Depending upon the design, by the time the final round is completed, the study will have an overall response rate of between 31 and 35 percent. This is assuming that respondents who miss a round will continue to be included in future rounds of data collection. If this is not the case the response rates would drop even further. While weighting can potentially account for some nonresponse bias, given the limitation to the weighting methods, it becomes risky to depend on weighting to compensate for nonresponse rates of approximately 70 percent.

For the youth cohort, an initial sample size of over 10,000 persons will be needed to yield 5,000 final respondents for Scenarios B and C. If this sample were doubled (so that estimates can be made for subgroups such as males versus females or 12 to 15 year olds versus 16 to 17 year olds), nearly every 12- to 17-year-old respondent from the NSDUH would be sampled. As a result, high-risk youths could no longer be oversampled, and the design would be less efficient. For this reason, if the youth sample were to include about 20,000 respondents, the recruitment for the longitudinal study might need to take place over 2 survey years. This may also be affected by the planned NSDUH survey update, where a proposal to reduce the number of 12 to 25 year olds sampled every year is under consideration.

Table 16. Adult Cohort: Assumed Response Rates and Sample Sizes for Longitudinal Design Scenarios A, B, and C

Study/Phase	Measure	Scenario A	Scenario B	Scenario C
NSDUH	*Screening Response Rate*	*0.89*	*0.89*	*0.89*
	Interview Response Rate	*0.74*	*0.74*	*0.74*
	NSDUH Sample (18+)	45,000	45,000	45,000
	Eligibility Rate - English Only	*0.96*	*0.96*	*0.96*
Longitudinal Study	Selected Sample	975	9,754	10,441
Baseline	*Assumed Response Rate*	*0.67*	*0.67*	*0.67*
	Expected Sample Size	658	6,579	7,042
	Past Year Prevalence of Any Mental Illness	*0.19*	*0.19*	*0.19*
6-Month 1	Eligible for 6-Month 1	NA	1,232	1,318
	Assumed Response Rate	*NA*	*0.86*	*0.86*
	Expected Sample Size	NA	1,059	1,134
Annual 1	*Assumed Response Rate*	*0.81*	*0.81*	*0.81*
	Expected Sample Size	533	5,329	5,704
Annual 2	*Assumed Response Rate*	*0.76*	*0.76*	*0.76*
	Expected Sample Size	500	5,000	5,352
Long Term	*Assumed Response Rate*	*NA*	*NA*	*0.71*
	Expected Sample Size	NA	NA	5,000
	Overall Response Rate	*0.33*	*0.33*	*0.31*

NA = Not applicable.

Table 17. Youth Cohort: Assumed Response Rates and Sample Sizes for Longitudinal Design Scenarios A, B, and C

Study/Phase	Measure	Scenario A	Scenario B	Scenario C
NSDUH	*Screening Response Rate*	*0.89*	*0.89*	*0.89*
	Interview Response Rate	*0.85*	*0.85*	*0.85*
	NSDUH Sample (12–17)	22,500	22,500	22,500
	Eligibility Rate—English Only	*0.96*	*0.96*	*0.96*
Longitudinal Study	Selected Sample	1,073	10,725	11,525
Baseline	*Assumed Response Rate*	*0.65*	*0.65*	*0.65*
	Expected Sample Size	694	6,944	7,463
	Past Year Prevalence of Any Mental Illness	*0.10*	*0.10*	*0.10*
6-Month 1	Eligible for 6-Month 1	NA	666	716
	Assumed Response Rate	*NA*	*0.82*	*0.82*
	Expected Sample Size	NA	546	587
Annual 1	*Assumed Response Rate*	*0.77*	*0.77*	*0.77*
	Expected Sample Size	535	5,347	5,746
	Past Year Incidence of Any Mental Illness	*0.09*	*0.09*	*0.09*
6-Month 2	Eligible for 6-Month 2	NA	1,145	1,231
	Assumed Response Rate	*NA*	*0.75*	*0.75*
	Expected Sample Size	NA	853	917
Annual 2	*Assumed Response Rate*	*0.72*	*0.72*	*0.72*
	Expected Sample Size	500	5,000	5,373
Long Term	*Assumed Response Rate*	*NA*	*NA*	*0.67*
	Expected Sample Size	NA	NA	5,000
	Overall Response Rate	*0.35*	*0.35*	*0.33*

NA = Not applicable.

4.4 Cost Estimates for Design Scenarios

Table 18 provides a summary of the three mental health longitudinal design scenarios, including the advantages and disadvantages of each design. As shown in the table, cost estimates have been provided for a total of six design options, including two instrument options for each design scenario. In addition to considering costs associated with management; instrumentation; data collection preparations and implementation; data programming , management, and processing; analyses and reporting; and data file documentation and dissemination, the cost estimates include the following key design assumptions:

- **NSDUH Questionnaire Time**: A 5-minute SDQ will be added to the NSDUH main questionnaire to be completed by 12 to 17 year olds. The subsample for the longitudinal component will be recruited at the end of the main study interview.
- **Diagnostic Interview Time**: The SCID for adults and K-SADS for youth will each average 90 minutes. The adult and youth versions of the CIDI will average 120 minutes. Parent interviews for both instrument options will average 45 minutes. The 6-month adult and youth interviews will average 30 minutes, and the 6-month parent interview will average 20 minutes.
- **Interviewer Type**: Similar to the SCID procedures on the MHSS, clinical interviewers will administer paper-and-pencil versions of the SCID and K-SADS over the phone out of their home. The interviews will also be recorded. The completed SCID and K-SADS interviews will be shipped in-house for clinical and technical editing and keying. An electronic version of the CIDI will be administered by bachelor-level interviewers over the phone in a centralized call center environment. The cost estimates are based on current interviewer pay rates.
- **Incentive**: Based on the length of the instruments, adult and youth respondents who complete the baseline SCID and K-SADS, respectively, will receive $50. Adult and youth who complete the baseline CIDI will receive $60. Parent respondents will receive a baseline incentive of $30. The incentive for all 6-month assessments will be $15. The incentives will increase by $10 each year.
- **Tracing**: Advance mailings, batch tracing, and interactive tracing will be conducted for each follow-up period. Limited field tracing has been assumed as a last resort for cases with a potential address lead that do not have a phone number. For locating purposes only, it is also assumed that SSNs will be collected when possible.

The cost of a pilot study was not included in each design scenario estimate, but a small pilot study of approximately 200 cases would be needed to evaluate the selection methods, instrument, materials, procedures, and respondent burden to determine if improvements need to be made to meet the objectives of the study. If a pilot was conducted, it would take two years of preparation to implement the longitudinal study to allow time for selecting and programming the instruments and/or systems, developing training and field materials, conducting cognitive and usability testing, completing data collection preparations and training, and conducting the pilot and implementing any resulting improvements.

To demonstrate the distribution of costs throughout the mental health longitudinal study, Table 19 provides the cost estimate by year for Scenarios A, B, and C for the SCID/K-SADS instrument option. The CIDI instrument option would follow a similar distribution. The cost estimate assumes that preparations will occur in year 1 and the baseline study will begin in year 2. Due to NSDUH's quarterly design, data collection for the longitudinal study will be mostly continuous with the exception of the adult long-term follow-up for Scenario C. This continuous design will help keep costs to a minimum. In the one year when data collection will not occur between the youth long-term follow-up and the adult long-term follow-up, there will be close-out activities and analysis and reporting associated with the youth sample and tracing, data collection preparations, and training for the adult long-term follow-up.

If there is an interest in increasing the sample size in order to yield subgroup estimates, Figure 2 provides a marginal cost curve using the SCID/K-SADS option for Scenario A as an example. The base sample size for Scenario A is presented in Tables 16 and 17. It assumes a final sample size of 500 completed interviews each for the adult and youth samples, for a total of 1,000 completed interviews (and 500 parent pair completes). Figure 2 shows the estimated percentage cost increase over the base price for increasing the base sample by 500 adult and youth interviews at a time. For example, if the final sample is increased to 2,000 adult interviews and 2,000 youth interviews (and 2,000 parent pair interviews), the cost will increase by approximately 173 percent of the base price.

Figure 2. Marginal Cost Curve Using the SCID/K-SADS Option for Scenario A

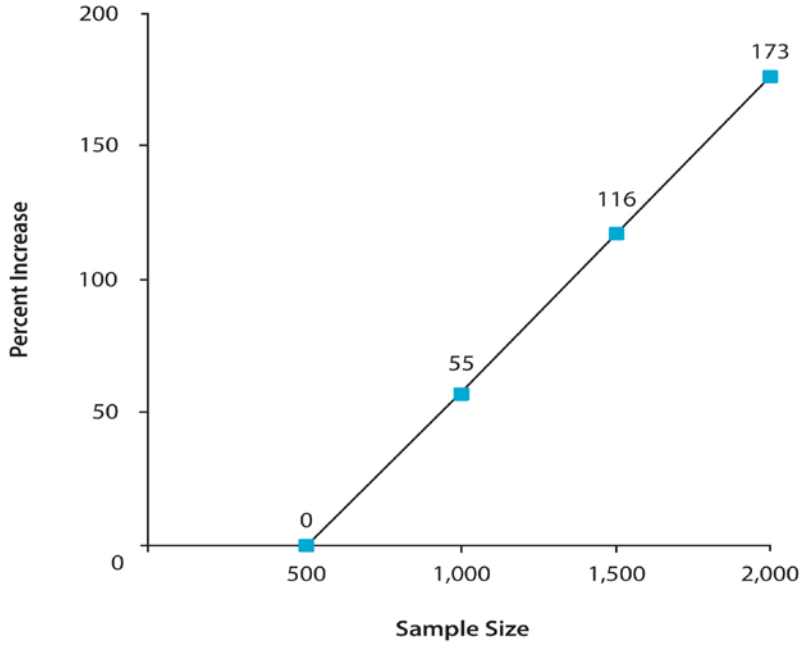


Table 18. Cost Estimates for Longitudinal Design Scenarios

Design Scenario	General Description	Advantages	Disadvantages	Data Collection Period	Waves	Baseline Interviews	Instrument	Cost Estimate
A	Baseline plus two annual assessments	• Prevalence and incidence • Development of comorbidity • Least burdensome option for participants	• No course data unless sample sizes are increased	3 years	3	1,352 (plus 694 parent interviews)	CIDI for 12 or older	
							SCID for 18 or older; K-SADS for 12-17	
B	Baseline plus 6-month assessments and two annual assessments	• Annual advantages mentioned for Scenario A • Captures course for episodic short-term disorders • Captures data on related factors (e.g., extent of service use) that may be transient or of short duration	• Limited long-term course data • Limited remission and recurrence data	3 years	4 to 5, depending on the age of the respondent at baseline	13,523 (plus 6,944 parent interviews)	CIDI for 12 or older	
							SCID for 18 or older; K-SADS for 12-17	
C	Baseline plus 6-month assessments, two annual follow-ups, and long-term follow-ups	• Annual and 6-month advantages mentioned for Scenarios A and B • Long-term course data (remission and recurrence)	• Most burdensome option for participants	4 to 7 years, depending on age of respondent at baseline	5 to 6, depending on the age of the respondent	14,506 (plus 7,463 parent interviews)	CIDI for 12 or older	
							SCID for 18 or older; K-SADS for 12–17	

CIDI = Composite International Diagnostic Interview; K-SADS = Kiddie-Schedule of Affective Disorders and Schizophrenia; SCID = Structured Clinical Interview for the DSM-IV.

Table 19. Cost Estimate by Year for Scenarios A, B, and C for the SCID/K-SADS Instrument Option

Year	Scenario A		Scenario B		Scenario C	
	Estimated Cost	Percent	Estimated Cost	Percent	Estimated Cost	Percent
1		16%		4%		3%
2		31%		37%		30%
3		24%		30%		24%
4		23%		27%		21%
5		6%		2%		12%
6		N/A		N/A		1%
7		N/A		N/A		7%
8		N/A		N/A		1%
Total		**100%**		**100%**		**100%**

5. Conclusion

The main analytic goal of a mental health longitudinal component to NSDUH would be to understand the trajectories of mental health problems. These include onset, persistence, remission, and relapse, as well as comorbidity and predictors related to mental health problems.

Various mental health longitudinal study designs would permit estimation of the prevalence, incidence (new cases), or course (e.g., duration, number of episodes) of mental disorders. When developing the mental health longitudinal design scenarios, RTI weighed the analytic goals of the study against the feasibility of conducting such a study and the impact of the design on the main NSDUH and costs. Literature on estimated prevalence and incidence rates and life course of mental disorders was also researched to inform the number and length of waves and sample size. In addition, RTI examined the various options and issues for consideration, such as the measures of interest, instrumentation, analytic capabilities, sample issues, and data collection implications, as well the potential impact to various main study operations.

This report, design scenarios, and cost estimates should provide SAMHSA and NIMH with the background needed to assess the feasibility of including a mental health longitudinal component in the NSDUH, including the potential effect of a longitudinal component on the main study and related issues for consideration. This report, however, does not address additional consequences of expanding the NSDUH to include a longitudinal component, such as the staffing needs or contractual consequences from the federal perspective.

6. References

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40*(1), 57-87. doi: 10.1111/1469-7610.00424

Batelaan, N. M., de Graaf, R., Spijker, J., Smit, J. H., van Balkom, A. J., Vollebergh, W. A., & Beekman, A. T. (2010). The course of panic attacks in individuals with panic disorder and subthreshold panic disorder: a population-based study. *Journal of Affective Disorders, 121*(1-2), 30-38. doi: S0165-0327(09)00186-4 [pii] 10.1016/j.jad.2009.05.003

Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development, 24*(2), 145-152.

Bird, H. R., Shaffer, D., Fisher, P. W. et al. (1993). The Columbia Impairment Scale (CIS): Pilot findings on a measure of global impairment for children and adolescents. *International Journal of Methods in Psychiatric Research, 3*, 167-176.

Bourdon, K. H., Goodman, R., Rae, D. S., Simpson, G., & Koretz, D. S. (2005). The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 557-564.

Brick, J. M., Edwards, W. S., & Lee, S. (2007). Sampling telephone numbers and adults, interview length, and weighting in the California Health Interview Survey Cell Phone Pilot Study. *Public Opinion Quarterly*, *71*(5), 793-813.

Cantor, D., O'Hare, B. C., & O'Connor, K. S. (2007). The use of monetary incentives to reduce nonresponse in random digit dial telephone surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.) *Advances in telephone survey methodology* (pp. 471-498). Hoboken, NJ: Wiley.

Castiglioni, L., Pffor, K., & Krieger, U. (2008). The effect of incentives on response rates and panel attrition: Results of a controlled experiment. Survey Research Methods, 2, 151-158.

Copeland, W., Shanahan, L., Costello, E. J., & Angold, A. (2011). Cumulative prevalence of psychiatric disorders by young adulthood: A prospective cohort analysis from the Great Smoky Mountains Study. *Journal of the American Academy of Child and Adolescent Psychiatry, 50*(3), 252-261.

Costello, E. J., Angold, A., Burns, B. J., Stangl, D. K., Tweed, D. L., Erkanli, A., et al. (1996). The Great Smoky Mountains Study of Youth. Goals, design, methods, and the prevalence of DSM-III-R disorders. *Archives of General Psychiatry, 53*(12), 1129-1136.

Costello, E. J., Foley, D. L., & Angold, A. (2006). 10-year research update review: The epidemiology of child and adolescent psychiatric disorders: II. Developmental epidemiology. *Journal of the American Academy of Child & Adolescent Psychiatry, 45*(1), 8-25. doi: 10.1097/01.chi.0000184929.41423.c0

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry, 60*(8), 837-844. doi: 10.1001/archpsyc.60.8.837 60/8/837 [pii]

Coughlin, S. S. (1990). Recall bias in epidemiologic studies. *Journal of Clinical Epidemiology, 43*(1), 87-91.

Couper, M. P., & Ofstedal, M. B. (2009). Keeping in contact with mobile sample members. In P. Lynn (Ed.) *Methodology of longitudinal surveys* (pp. 183-203). Chichester, UK: John Wiley & Sons, Inc.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.

Eaton, W. W., Kalaydjian, A., Scharfstein, D. O., Mezuk, B., & Ding, Y. (2007). Prevalence and incidence of depressive disorder: The Baltimore ECA follow-up, 1981-2004. *Acta Psychiatrica Scandinavica, 116*(3), 182-188. doi: ACP1017 [pii] 10.1111/j.1600-0447.2007.01017.x

Ebner-Priemer, U. W., Kuo, J., Welch, S. S., Thielgen, T., Witte, S., Bohus, M., & Linehan, M. M. (2006). A valence-dependent group-specific recall bias of retrospective self-reports: A study of borderline personality disorder in everyday life. *The Journal of Nervous and Mental Disease, 194*(10), 774-779. doi: 10.1097/01.nmd.0000239900.46595.72 00005053-200610000-00008 [pii]

Elliott, D. S., Huizinga, D., & Menard, S. (1989). Multiple problem youth: Delinquency, substance use, and mental health problems. Research in criminology. New York, NY: US: Springer-Verlag Publishing.

Fennig, S., Craig, T., Lavelle, J., Kovasznay, B., & Bromet, E. J. (1994) Best-estimate versus structured interview-based diagnosis in first-admission psychosis. *Comprehensive Psychiatry*, *35*, 341-348.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR axis 1 disorders, research version, non-patient edition*. New York, NY: Biometrics Research, New York State Psychiatric Institute.

French, D. P., & Sutton, S. (2010). Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *British Journal of Health Psychology*, *15*, 453-468.

Gelman, A., Stevens, M., & Chan, V. (2003). Regression modeling and meta-analysis for decision making: A cost-benefit analysis of incentives in telephone surveys. *Journal of Business & Economic Statistics*, *21*(2), 213-225.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581-586.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry*, *177*, 534-539. Retrieved from http://bjp.rcpsych.org/cgi/content/full/177/6/534

Grant, B. F., & Kaplan, K. D. (2005). *Source and accuracy statement for the wave 2 national epidemiologic survey on alcohol and related conditions (NESARC)*. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Grant, B. F., Dawson, D. A., Stinson, F. S., Chou, P. S., Kay, W., & Pickering, R. (2003). The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): Reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. *Drug and Alcohol Dependence, 71*(1), 7-16.

Grant, B. F., Kaplan, K., Shepard, J., & Moore, T. (2003). Source and accuracy statement for Wave 1 of the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Haggerty, K., Fleming, C., Catalano, R., Petrie, R., Rubin, R., & Grassley, M. (2007). Ten years later: Locating and interviewing children of drug abusers. *Evaluation and Program Planning, 31*, 1-9.

Happ, M. B., Sereika, S., Garrett, K., & Tate, J. (2008). Use of the quasi-experimental sequential cohort design in the Study of Patient-Nurse Effectiveness with Assisted Communication Strategies (SPEACS). *Contemporary Clinical Trials, 29*(5), 801-808. doi: S1551-7144(08)00079-7 [pii] 10.1016/j.cct.2008.05.010

Hudson, J. I., Hiripi, E., Pope, H. G., Jr., & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological Psychiatry, 61*(3), 348-358. doi: S0006-3223(06)00474-4 [pii] 10.1016/j.biopsych.2006.03.040

Hudson, J. I., Hiripi, E., Pope, H. G., Jr., & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological Psychiatry, 61*(3), 348-358.

Jackson, J. S., Torres, M., Caldwell, C. H., Neighbors, H. W., Nesse, R. M., Taylor, R. J., et al. (2004). The National Survey of American Life: A study of racial, ethnic and cultural influences on mental disorders and mental health. *International Journal of Methods in Psychiatric Research*, *13*, 196–207.

Jäckle, A. (2008). The causes of seam effects in panel surveys. Retrieved from http://www.iser.essex.ac.uk/publications/working-papers/iser/2008-14.pdf

Jensen, P. S., Rubio-Stipec, M., Canino, G., Bird, H. R., Dulcan, M. K., Schwab-Stone, M. E., et al. (1999). Parent and child contributions to diagnosis of mental disorder: Are both informants always necessary? *Child and Adolescent Psychiatry*, *38*(12), 1569–1579.

Jewell, J., Handwerk, M., Almquist, J., & Lucas, C. (2004). Comparing the validity of clinician-generated diagnosis of conduct disorder to the diagnostic interview schedule for children. *Journal of Clinical Child & Adolescent Psychology*, 33, 536-546.

Johnson, E. O., & Schultz, L. (2005 September). Forward telescoping bias in reported age of onset: An example from cigarette smoking. *International Journal of Methods in Psychiatric Research*, *14*(3), 119-129.

Jordan, B. K., Karg, R. S., Batts, K. R., Wiesen, C. A., & Epstein, J. F. (2008). A clinical validation of the National Survey on Drug Use and Health Assessment of Substance Use Disorders. *Addictive Behaviors*, *33*, 782-798.

Kessler, R. C., & Merikangas, K. R. (2004). The National Comorbidity Survey Replication (NCS-R): background and aims. *International Journal of Methods in Psychiatric Research, 13*(2), 60-68.

Kessler, R. C., & Wang, P. S. (2008). The descriptive epidemiology of commonly occurring mental disorders in the United States. *Annual Review of Public Health, 29*(1), 115-129.

Kessler, R. C., Avenevoli, S., Green, J., Gruber, M. J., Guyer, M., He, Y., et al. (2009, April). National Comorbidity Survey Replication Adolescent Supplement (NCS-A): III. Concordance of DSM-IV/CIDI diagnoses with clinical reassessments. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(4), 386-399. doi: 10.1097/CHI.0b013e31819a1cbc

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 593-602. doi: 62/6/593 [pii] 10.1001/archpsyc.62.6.593

Kessler, R. C., Zhao, S., Blazer, D. G., & Swartz, M. (1997). Prevalence, correlates, and course of minor depression and major depression in the National Comorbidity Survey. *Journal of Affective Disorders, 45*(1-2), 19-30. doi: S0165-0327(97)00056-6 [pii]

Kessler, R. C., Berglund, P., Chiu, W. T., Demler, O., Heeringa, S., Hiripi, E., et al. (2004). The U.S. National Comorbidity Survey Replication (NCS-R): Design and field procedures. *International Journal of Methods in Psychiatric Research*, *13*(2) 69-92.

Kranzler, H. R., Kadden, R. M., Babor, T. F., Tennen, H., & Rounsaville, B. J. (1996) Validity of the SCID in substance abuse patients. *Addiction*, *91*, 859-868.

Laurie, H., & Lynn, P. (2009). The use of respondent incentives on longitudinal surveys. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 205-234). London: John Wiley and Sons.

Lee, L., & Carr, J. (2009). *Evaluation of a 12-month reference period in the National Crime Victimization Survey (NCVS)*. NORC at the University of Chicago. Retrieved from http://www.fcsm.gov/09papers/Lee_II-A.pdf

Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *Journal of Abnormal Psychology, 102*(1), 133-144.

Longwell, B. T., & Truax, P. (2005). The differential effects of weekly, monthly, and bimonthly administrations of the Beck Depression Inventory-II: Psychometric properties and clinical implications. *Behavior Therapy*, *36*, 265-275.

Lyneham, H. J., & Rapee, R. M. (2005). Agreement between telephone and in-person delivery of a structured interview for anxiety disorders in children. *Journal of the American Academy of Child & Adolescent Psychiatry, 44*(3), 274-282.

Meehan, A. E., Saleska, E. L,. Hinsdale-Shouse, M. A., Kinsey, N. L., & Tischner, E. C. (2009, May). *The challenges of locating young adults for a longitudinal study: Improved tracing strategies implemented for the National Longitudinal Study of Adolescent Health*. Presented at American Association for Public Opinion Research, Hollywood, FL.

Merikangas, K. R., He, J. P., Brody, D., Fisher, P. W., Bourdon, K., Koretz, D. S. (2010). Prevalence and treatment of mental disorders among U.S. children in the 2001-2004 NHANES. *Pediatrics*, *125*(1), 75-81.

Merikangas, K. R., Jin, R., He, J. P., Kessler, R. C., Lee, S., Sampson, N. A., et al. (2011). Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of General Psychiatry, 68*(3), 241-251. doi: 68/3/241 [pii] 10.1001/archgenpsychiatry.2011.12

Millstein, R. B., Wilens, T. E., Biederman, J., & Spencer, T. J. (1997). Presenting ADHD symptoms and subtypes in clinically referred adults with ADHD. *Journal of Attention Disorders, 2*(3), 159-166. doi: 10.1177/108705479700200302

Mineka, S., & Nugent, K. (1995). Mood-congruent memory biases in anxiety and depression. In D. Schacter (Ed.), *Memory distortion: How minds, brains and societies reconstruct the past* (pp. 173–193). Cambridge, Massachusetts: Harvard University Press.

Miyazaki, Y., & Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods, 5*(1), 44-63.

Neter, J. & Waksberg, J. (1964). A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association*, 59, 18-55.

NIH/National Institute on Alcohol Abuse and Alcoholism. (October 4, 2005). National survey sharpens picture of major depression among U.S. adults. *ScienceDaily*. Retrieved from http://www.sciencedaily.com/releases/2005/10/051004084117.htm

Paing, W. W., Weller, R. A., Dixon, T. A., & Weller, E. B. (2010). Face-to-face versus telephone administration of the Parent's Version of the Children's Interview for Psychiatric Syndromes (P-ChIPS). *Current Psychiatry Reports*, *12*(2), 122-126.

Pierret, C. R. (2001). Event history data survey recall: An analysis of the National Longitudinal Survey of Youth 1979 recall experiment. Journal of Human Resources, 36(3), 439-466.

Perala, J., Suvisaari, J., Saarni, S. I., Kuoppasalmi, K., Isometsa, E., Pirkola, S., et al. (2007). Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Archives of General Psychiatry, 64*(1), 19-28. doi: 64/1/19 [pii] 10.1001/archpsyc.64.1.19

Prusoff, B. A., Merikangas, K. R., & Weissman, M. M. (1988). Lifetime prevalence and age of onset of psychiatric disorders: Recall 4 years later. *Journal of Psychiatric Research*, *22(2)*, 107-117.

Ramirez Basco, M., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendrickse, W., & Barnett, V. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *American Journal of Psychiatry*, *157*, 1599-1605.

Raudenbush, S. W., & Chan, W. S. (1993). Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of Consulting and Clinical Psychology, 61*(6), 941-951.

Regier, D. A., Myers, J. K., Kramer, M., Robins, L. N., Blazer, D. G., Hough, R. L., et al. (1984). The NIMH Epidemiologic Catchment Area Program: Historical context, major objectives, and study population characteristics. *Archives of General Psychiatry, 41*(10), 934-941.

Ringel, J. S., & Sturm, R. (2001). National estimates of mental health utilization and expenditures for children in 1998. *Journal of Behavioral Health Services Research, 28*(3), 319-333.

Rips, L. J., Conrad, F. G., & Fricker, S. S. (2003). Straightening the seam effect in panel surveys. *Public Opinion Quarterly*, *67*, 522-554.

Robins, L. N., & Cottler, L. B. (2004). Making a structured psychiatric diagnostic interview faithful to the nomenclature. *American Journal of Epidemiology, 160*(8), 808-813.

Robins, L. N., Helzer, J. E., Croughan, J., & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Archives of General Psychiatry, 38*(4), 381-389.

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke, J., et al. (1988). The Composite International Diagnostic Interview: An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry, 45*(12), 1069-1077.

Rodgers, W. L. (2011). Effects of increasing the incentive size in a longitudinal study. *Journal of Official Statistics*, *27*(2), 279-299.

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1997). Comparability of telephone and face-to-face interviews in assessing axis I and axis II disorders. *American Journal of Psychiatry*, *154*, 1593-1598.

Rowe, R., Costello, E. J., Angold, A., Copeland, W. E., & Maughan, B. (2010). Developmental pathways in oppositional defiant disorder and conduct disorder. *Journal of Abnormal Psychology, 119*(4), 726-738. doi: 10.1037/a0020798

Ruan, W. J., Goldstein, R. B., Chou, S. P., Smith, S. M., Saha, T. D., Pickering, R. P., et al. (2008). The Alcohol Use Disorder and Associated Disabilities Interview Schedule IV (AUDADIS-IV): Reliability of new psychiatric diagnostic modules and risk factors in a general population sample. *Drug and Alcohol Dependence*, *92*, 27-36.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*(5328), 918-924.

Segal, D. L., Kabacoff, R. I., Hersen, M., Van Hasselt, V. B., & Ryan, C. F. (1995). Update on the reliability of diagnosis in older psychiatric outpatients using the structured clinical interview for DSM-III-R. *Journal of Clinical Geropsychology*, *1*, 313-321.

Shaffer, D., Gould, M. S., Brasic, J., Ambrosini P, Fisher P, Bird H, Aluwahlia S (1983). A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry 40*,1228-1231.

Shear, M. K., Greeno, C., Kang, J., Ludewig, D., Frank, E., Swartz, H. A., & Hanekamp, M. (2000). Diagnosis of nonpsychotic patients in community clinics. *American Journal of Psychiatry*, *157*, 581-587.

Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. In R.M. Groves, D.A. Dillman, J.L. Eltinge, & R.J.A. Little (Eds.) *Survey Nonresponse* (pp. 163-177). New York: Wiley.

Singer, E., & Bossarte, R. M. (2006). Incentives for survey participation when are they "coercive"? *American Journal of Preventive Medicine, 31*(5), 411-418.

Skoog, G., & Skoog, I. (1999). A 40-year follow-up of patients with obsessive-compulsive disorder [see comments]. *Archives of General Psychiatry, 56*(2), 121-127.

Sobin, E. Weissman, M. M., Goldstein, R. B., Adams, P. (1993). Diagnostic interviewing for family studies: Comparing telephone and face-top-face methods for the diagnosis of lifetime psychiatric disorders. *Psychiatric Genetics*, *3*, 227-233.

Solomon, D. A., Leon, A. C., Coryell, W. H., Endicott, J., Li, C., Fiedorowicz, J. G., et al. (2010). Longitudinal course of bipolar I disorder: Duration of mood episodes. *Archives of General Psychiatry, 67*(4), 339-347. doi: 67/4/339 [pii] 10.1001/archgenpsychiatry.2010.15

Steiner, J. L., Tebes, J. K., Sledge, W. H., & Walker, M. L. (1995) A comparison of the structured clinical interview for DSM-III-R and clinical diagnoses. *Journal of Nervous and Mental Disease*, *183*, 365-369.

The tenth mental measurements yearbook. (1991). O. K. Buros (Series Ed.). Lincoln, NE: Buros Institute of Mental Measurements.

Tiihonen, J., Isohanni, M., Rasanen, P., Koiranen, M., & Moring, J. (1997). Specific major mental disorders and criminality: A 26-year prospective study of the 1966 northern Finland birth cohort. *American Journal of Psychiatry, 154*(6), 840-845.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Uhrig, S., & Sala, E. (2011) When change matters: An analysis of survey interaction in dependent interviewing on the British household panel study. *Sociological Methods and Research*, *40*(333).

Warren, J. R., & Halpern-Manners, A. (2011). Panel conditioning in longitudinal sociological surveys. Retrieved from http://www.soc.umn.edu/~warre046/Warren%20&%20Halpern-Manners%20PC.pdf

Wilson, S. & Howell, B. L. (2007). Disease prevalence and survey design effects: A response to Weir and Smith. *Social Science and Medicine*, 65, 1078-1081.

Wilson, S. E., & Howell, B. L. (2005). Do panel surveys make people sick? U.S. arthritis trends in the Health and Retirement Study. *Social Science and Medicine*, 60, 2623-2627.

Wittchen, H. U., Gloster, A. T., Beesdo-Baum, K., Fava, G. A., & Craske, M. G. (2010). Agoraphobia: A review of the diagnostic classificatory position and criteria. *Depress Anxiety, 27*(2), 113-133. doi: 10.1002/da.20646

Zanarini, M. C., & Frankenburg, F. R. (2001). Attainment and maintenance of reliability of axis I and axis II disorders over the course of a longitudinal study. *Comprehensive Psychiatry*, *42*(5), 369-374.

Zanarini, M. C., Skodol, A. E., Bender, D., Dolan, R., Sanislow, C., Schaefer, E., et al. (2000). The Collaborative Longitudinal Personality Disorders Study: Reliability of axis I and II diagnoses. *Journal of Personality Disorders*, *14*(4), 291-299

Vicente, N.-A., & Woodworth, G. G. (1994). Analysis of longitudinal data with unequally spaced observations and time- dependent correlated errors. *Biometrics*, *50*(2), 455-456.

Thwin, S. S., & Gagnon, C. (2008). Longitudinal data observed over unequal time intervals: Modeling correlations in the analysis of binary outcomes data. KG, Germany: VDM Verlag Dr. Müller Aktiengesellschaft Co.

Zimmerman, M., & Mattia, J. I. (2001). The Psychiatric Diagnostic Screening Questionnaire: Development, reliability and validity. *Comprehensive Psychiatry*, *42*, 175-189.

Appendix A:
Longitudinal Design Scenario Diagrams

This page intentionally left blank

Figure A-1. NSDUH Mental Health Longitudinal Study Design: Scenario A

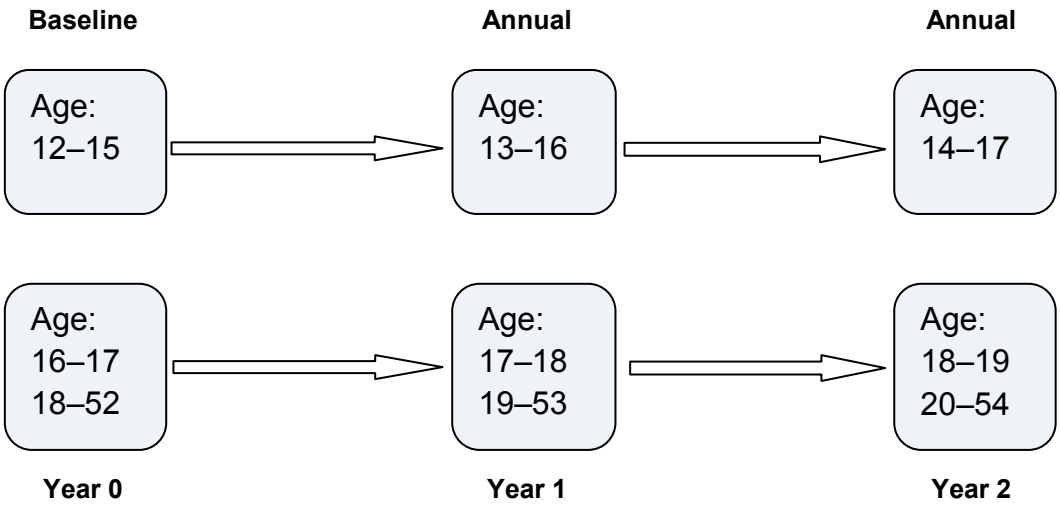


Figure A-2. NSDUH Mental Health Longitudinal Study Design: Scenario B

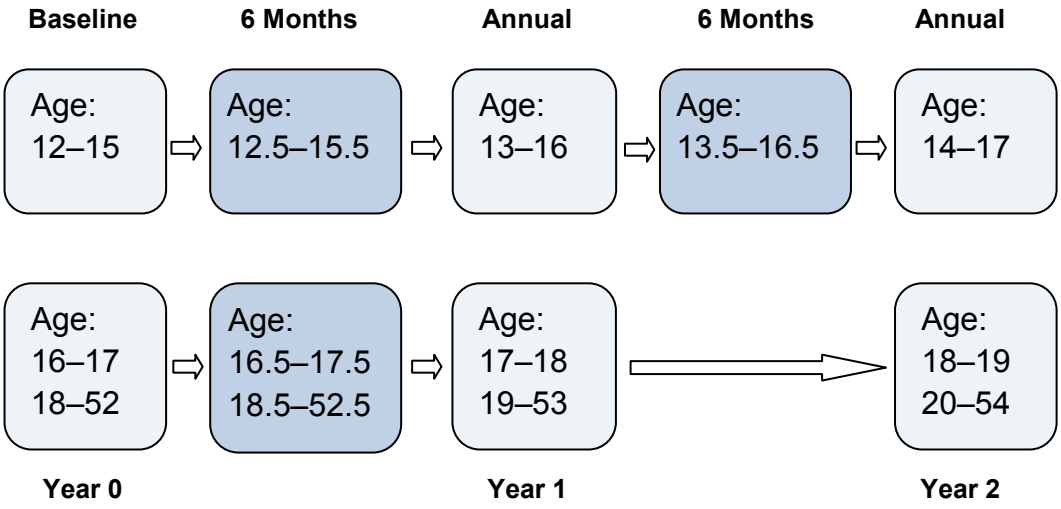
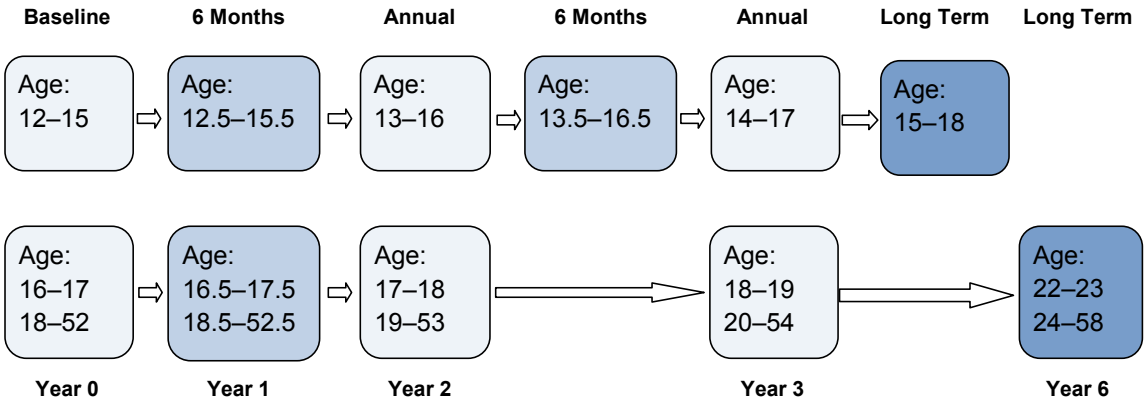



Figure A-3. NSDUH Mental Health Longitudinal Study Design: Scenario C