

# **Reliability of Key Measures in the National Survey on Drug Use and Health**

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Office of Applied Studies

## **Acknowledgments**

This report was prepared by the Office of Applied Studies (OAS), Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services (HHS), and by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina. Work by RTI was performed under Contract No. 283-2004-00022. The coauthors at SAMHSA and RTI in alphabetical order are James R. Chromy, Moshe Feder, Joseph C. Gfroerer, Erica Hirsch, Joel Kennet, Katherine B. Morton, Lanny Piper, Beth H. Riggsbee, Jeanne A. Snodgrass, Thomas G. Virag, and Feng Yu. Arthur Hughes was the Government Project Officer for NSDUH at OAS, SAMHSA.

## **Public Domain Notice**

All material appearing in this report is in the public domain and may be reproduced or copied without permission from SAMHSA. Citation of the source is appreciated. However, this publication may *not* be reproduced or distributed for a fee without the specific, written authorization of the Office of Communications, SAMHSA, HHS.

## **Recommended Citation**

Substance Abuse and Mental Health Services Administration. (2010). *Reliability of Key Measures in the National Survey on Drug Use and Health* (Office of Applied Studies, Methodology Series M-8, HHS Publication No. SMA 09-4425). Rockville, MD.

## **Electronic Access and Copies of Publication**

This publication may be downloaded from <http://www.oas.samhsa.gov>. Hard copies may be obtained from <http://www.oas.samhsa.gov/copies.cfm>. Or please call SAMHSA's Health Information Network at 1-877-SAMHSA-7 (1-877-726-4727) (English and Español).

## **Originating Office**

Substance Abuse and Mental Health Services Administration  
Office of Applied Studies  
Division of Population Surveys  
1 Choke Cherry Road, Room 7-1044  
Rockville, MD 20857

February 2010

# Table of Contents

Chapter	Page
List of Figures .....	vi
List of Tables .....	vii
Executive Summary .....	1
1. Introduction.....	5
1.1. Summary of NSDUH.....	5
1.2. Study Objectives .....	6
1.3. Content and Organization of This Report.....	7
2. Prior Reinterview Studies and Measures of Reliability .....	9
2.1. U.S. Decennial Census.....	11
2.2. Other Census Bureau Surveys .....	12
2.3. National Household Education Survey .....	12
2.4. Diagnostic and Validation Studies.....	13
3. Methods and Procedures .....	15
3.1. Overview of Reliability Study Procedures .....	15
3.1.1 Time between T1 and T2 Administrations .....	15
3.1.2 Same versus Different Interviewer Substudy .....	15
3.1.3 Incentive.....	16
3.2. Pretest.....	16
3.2.1 Pretest Methodology .....	16
3.2.2 Implications of the Pretest for the Full Study Implementation.....	18
3.3. Sample Design .....	20
3.4. Instrument Design .....	23
3.5. Administration of the Questionnaires .....	24
3.6. Refusals, Consent, and Incentives .....	26
3.7. Response Rates and Verification .....	27
3.8. Analysis Data .....	27
3.8.1 Types of Variables Examined.....	28
3.8.2 Agreement between T1 and T2 Responses.....	29
3.8.3 Interview Timing Data.....	29
4. Measures of Agreement and Approaches to Modeling.....	31
4.1. Measures of Agreement .....	31
4.1.1 Weighted Percent Reporting Consistently .....	31
4.1.2 Cohen's Kappa .....	32
4.1.3 Index of Inconsistency .....	34
4.2. Models.....	34
4.2.1 Response Propensity Models .....	34
4.2.2 Logit Model to Assess Dependence of Reliability of Age at First Use and Duration of Recall.....	35

## Table of Contents (continued)

Chapter	Page
4.2.3 Hui-Walter Modeling.....	35
5. Descriptive Survey Statistics .....	37
5.1. Survey Sample and Respondent Characteristics.....	37
5.2. Response Rates .....	38
5.3. Interview Timing Data.....	38
5.4. Time between T1 and T2 Interviews .....	40
6. Reliability Findings.....	43
6.1. Overview of Analyses.....	44
6.2. Reliability Statistics, by Type of Variable.....	44
6.2.1 Substance Use .....	45
6.2.2 Perceived Risk and Availability.....	46
6.2.3 Substance Dependence and Abuse.....	46
6.2.4 Illegal Behaviors and Marijuana Acquisition .....	47
6.2.5 Timing of Substance Use .....	47
6.2.6 Substance Use Treatment.....	47
6.2.7 Health Conditions and Health Care .....	48
6.2.8 Mental Health Treatment.....	48
6.2.9 Social Behavior and Attitudes .....	49
6.2.10 Mental Health Problems .....	49
6.2.11 Alcohol Use .....	50
6.2.12 Demographic Characteristics .....	50
6.3. Response Propensity Model Findings.....	51
6.3.1 First Interview Response Propensity Models .....	53
6.3.2 Reinterview Response Propensity Models.....	53
6.3.3 Comparison with the Full 2006 NSDUH Sample.....	53
6.4. Dependence of Age at First Use Reliability and Duration of Recall.....	53
6.4.1 General Observations on the Consistency of Age at First Use of Cigarettes, Alcohol, and Marijuana and Hashish and the Years Since First Use .....	54
6.4.2 Effect of the Duration of Recall on the Consistency of Age at First Use of Cigarettes, Alcohol, and Marijuana and Hashish and the Years Since First Use.....	54
6.4.3 Predicted Marginals and Their Interpretation.....	55
6.5. Hui-Walter Model and Findings.....	55
6.5.1 Hui-Walter Model.....	56
6.5.2 Results.....	56

## Table of Contents (continued)

<b>Chapter</b>	<b>Page</b>
7. Reliability Findings for Same versus Different Interviewer Substudy .....	117
8. Impact of Time between Interviews on Reliability.....	119
9. Recommendations for Future Reliability Studies .....	121
References.....	123
<b>Appendix</b>	
A Description of NSDUH.....	127
B Statistical Procedures .....	135
C Results of the T2 Interview Follow-Up Questions .....	139
D List of Contributors.....	151

## List of Figures

<b>Figure</b>	<b>Page</b>
5.1. Relationship between the T1 and T2 Durations of the Audio Computer-Assisted Self-Interviewing (ACASI) Interviews: 2006 NSDUH Reliability Study.....	41
B.1. Kappa Dependence on Prevalence Rates.....	135

## List of Tables

Table	Page
3.1. Pretest Response Rates, by Phase: 2006 NSDUH Reliability Study Pretest .....	19
3.2. Design Parameters at the Dwelling Unit Level: 2006 NSDUH Reliability Study .....	21
3.3. Design Parameters at the Person Level: 2006 NSDUH Reliability Study.....	22
3.4. Follow-Up Audio Computer-Assisted Self-Interviewing (ACASI) Questions in the T2 Interview: 2006 NSDUH Reliability Study.....	25
4.1. Basic 2x2 Contingency Table for Substance Use .....	31
5.1. Sample and Response Distribution: 2006 NSDUH Reliability Study .....	37
5.2. Final Unweighted and Weighted T1 and T2 Response Rates: 2006 NSDUH Reliability Study .....	39
5.3. Summary Statistics on Interview Length, in Minutes: 2006 NSDUH Reliability Study .....	39
5.4. Difference in Time to Complete Interview (in Minutes) at T1 and T2, with T2 by Same and Different FI, Including <i>p</i> Value for Test of Mean of T1 – T2 Being Zero: 2006 NSDUH Reliability Study.....	40
5.5. Correlations between T1 and T2 Timing Data: 2006 NSDUH Reliability Study .....	40
5.6. Distribution of Completed Cases by the Lag Time between T1 and T2 Audio Computer-Assisted Self-Interviewing (ACASI) Interviews in Days: 2006 NSDUH Reliability Study .....	41
6.1. Reliability Statistics for Substance Use in Lifetime among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	58
6.2. Reliability Statistics for Substance Use in the Past Year among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	59
6.3. Reliability Statistics for Age at First Use of Substances among Lifetime Users Aged 12 or Older: 2006 NSDUH Reliability Study .....	60
6.4. Reliability Statistics for First Substance Used among Lifetime Users Aged 12 or Older of Alcohol, Marijuana, or Cigarettes Who Reported the Same Age at First Use for at Least Two of the Substances: 2006 NSDUH Reliability Study.....	61
6.5. Reliability Statistics for Measures of Alcohol, Marijuana, and Cocaine Use among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	62

## List of Tables (continued)

Table	Page
6.6. Reliability Statistics for Blunt Use among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	63
6.7. Reliability Statistics for Lifetime Nonmedical Use of Specific Noncore Stimulants, Sedatives, Cough and Cold Medicines, or Hallucinogens and Past Year Nonmedical Use of Adderall® or Cough and Cold Medicines among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	64
6.8. Reliability Statistics for Perceived Risk and Availability of Substances, Participation in Risky Behaviors, and Likelihood of Smoking Cigarettes for Nonsmokers among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	65
6.9. Reliability Statistics for Alcohol Dependence or Abuse in the Past Year and the Criteria of Alcohol Dependence or Abuse among Past Year Alcohol Users Aged 12 or Older: 2006 NSDUH Reliability Study .....	67
6.10. Reliability Statistics for Marijuana Dependence or Abuse in the Past Year and the Criteria of Marijuana Dependence or Abuse among Past Year Marijuana Users Aged 12 or Older: 2006 NSDUH Reliability Study .....	68
6.11. Reliability Statistics for Cocaine Dependence or Abuse in the Past Year and the Criteria of Cocaine Dependence or Abuse among Past Year Cocaine Users Aged 12 or Older: 2006 NSDUH Reliability Study .....	69
6.12. Reliability Statistics for Pain Reliever Dependence or Abuse in the Past Year and the Criteria of Pain Reliever Dependence or Abuse among Past Year Pain Reliever Users Aged 12 or Older: 2006 NSDUH Reliability Study .....	70
6.13. Reliability Statistics for Substance Dependence or Abuse in the Past Year among Respective Substance Users Aged 12 or Older: 2006 NSDUH Reliability Study .....	71
6.14. Reliability Statistics for Encounters with the Police or the Court System, Driving under the Influence (DUI) of Alcohol or Illicit Drugs in the Past Year, and Legal Knowledge about Marijuana among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	72
6.15. Reliability Statistics for Marijuana Purchase and Acquisition among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	73
6.16. Reliability Statistics for Substance Use in the Year Prior to the Past Year among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	74



## List of Tables (continued)

<b>Table</b>	<b>Page</b>
6.17. Reliability Statistics for Age at Last Use of Substances among Lifetime Users Aged 12 or Older Who Last Used Respective Substance More Than 30 Days Ago: 2006 NSDUH Reliability Study .....	75
6.18. Reliability Statistics for Number of Months from Last Substance Use to Time of Initial Interview among Recent Substance Initiates Aged 12 or Older: 2006 NSDUH Reliability Study .....	76
6.19. Reliability Statistics for Substance Use Treatment among Substance Users Aged 12 or Older: 2006 NSDUH Reliability Study .....	77
6.20. Reliability Statistics for Past Year Health Care and Health Conditions among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	78
6.21. Reliability Statistics for Frequency of Past Year Health Care among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	79
6.22. Reliability Statistics for Lifetime Health Condition Variables among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	80
6.23. Reliability Statistics for Mental Health Treatment in the Past Year among Persons Aged 18 or Older: 2006 NSDUH Reliability Study .....	81
6.24. Reliability Statistics for Unmet Need for Mental Health Treatment/Counseling in the Past Year among Persons Aged 18 or Older: 2006 NSDUH Reliability Study.....	82
6.25. Reliability Statistics for Alternative Mental Health Treatment, Counseling, or Support in the Past Year among Persons Aged 18 or Older: 2006 NSDUH Reliability Study .....	83
6.26. Reliability Statistics for Social Environment Variables among Persons Aged 18 or Older: 2006 NSDUH Reliability Study .....	84
6.27. Reliability Statistics for School Experiences and Religious Beliefs among Persons Aged 12 to 17: 2006 NSDUH Reliability Study.....	85
6.28. Reliability Statistics for Perceptions of Cigarette, Alcohol, and Marijuana Use among Persons Aged 12 to 17: 2006 NSDUH Reliability Study .....	86
6.29. Reliability Statistics for Delinquent Behavior and Extracurricular Activities in the Past Year among Persons Aged 12 to 17: 2006 NSDUH Reliability Study.....	87

## List of Tables (continued)

<b>Table</b>	<b>Page</b>
6.30. Reliability Statistics for Serious Psychological Distress (SPD) and Components of SPD in the Past Year among Persons Aged 18 or Older: 2006 NSDUH Reliability Study .....	88
6.31. Reliability Statistics for Major Depressive Episode (MDE) and Methods of Treatment among Persons Aged 18 or Older: 2006 NSDUH Reliability Study .....	89
6.32. Reliability Statistics for Age at First Major Depressive Episode (MDE), Impact of Depression on Person's Life, and Amount Treatment Helped among Persons Aged 18 or Older: 2006 NSDUH Reliability Study .....	90
6.33. Reliability Statistics for Youth Mental Health Dichotomous Variables among Persons Aged 12 to 17: 2006 NSDUH Reliability Study .....	91
6.34. Reliability Statistics for Major Depressive Episode (MDE) and Methods of Treatment among Persons Aged 12 to 17: 2006 NSDUH Reliability Study .....	92
6.35. Reliability Statistics for Age at First Major Depressive Episode (MDE) and Impact of Depression on a Person's Life among Persons Aged 12 to 17 Who Were Classified as Having Lifetime MDE or Who Met the Suicide Ideation Criterion: 2006 NSDUH Reliability Study .....	93
6.36. Reliability Statistics for Alcohol Consumption among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	94
6.37. Reliability Statistics for Household Composition and Family Income among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	95
6.38. Reliability Statistics for Health Insurance Coverage among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	96
6.39. Reliability Statistics for Lack of Health Insurance Coverage among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	97
6.40. Reliability Statistics for Main Reason Stopped Being Covered by Health Insurance among Persons Aged 12 or Older Who Were Currently without Health Care Coverage But Had Health Care Coverage in the Past: 2006 NSDUH Reliability Study .....	98
6.41. Reliability Statistics for Types of Income Received among Persons Aged 12 or Older, by Income Sample: 2006 NSDUH Reliability Study .....	99
6.42. Reliability Statistics for Demographics (Dichotomized) among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	100

## List of Tables (continued)

Table	Page
6.43. Reliability Statistics for Demographic Scaled Variables among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	101
6.44. Reliability Statistics for Ethnicity, Race, and Marital Status (Dichotomized) among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	102
6.45. Reliability Statistics for Noncore Demographic Categorical Variables among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	103
6.46. Reliability Statistics for Industry and Occupation Codes among Persons Aged 15 or Older Who Were Ever Employed: 2006 NSDUH Reliability Study.....	104
6.47. Reliability Statistics for Nonmedical Use of Specific Pain Relievers in the Lifetime among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	105
6.48. Response Propensity Model for Completion of Initial Interview by Sample Members Aged 12 or Older: 2006 NSDUH Reliability Study .....	106
6.49. Response Propensity Model for Completion of Reinterviews by Sample Members Aged 12 or Older: 2006 NSDUH Reliability Study .....	107
6.50. Response Propensity Model for Completion of the NSDUH Interview among Main Study Sample Members Aged 12 or Older: 2006 NSDUH Reliability Study .....	108
6.51. Dependence of Age at First Use of Cigarettes Consistency and Duration of Recall for Persons Aged 18 or Older: 2006 NSDUH Reliability Study .....	109
6.52. Dependence of Age at First Use of Alcohol Consistency and Duration of Recall for Persons Aged 18 or Older: 2006 NSDUH Reliability Study .....	110
6.53. Dependence of Age at First Use of Marijuana and Hashish Consistency and Duration of Recall for Persons Aged 18 or Older: 2006 NSDUH Reliability Study .....	111
6.54. Dependence of Age at First Use of Cigarettes Consistency and Duration of Recall for Persons Aged 24 or Older: 2006 NSDUH Reliability Study .....	112
6.55. Dependence of Age at First Use of Alcohol Consistency and Duration of Recall for Persons Aged 24 or Older: 2006 NSDUH Reliability Study .....	113
6.56. Dependence of Age at First Use of Marijuana and Hashish Consistency and Duration of Recall for Persons Aged 24 or Older: 2006 NSDUH Reliability Study .....	114
6.57. Predicted Marginals for Proportion of Agreement between T1 and T2: 2006 NSDUH Reliability Study.....	115

## List of Tables (continued)

<b>Table</b>	<b>Page</b>
6.58. Estimated True Prevalence of Past Year Marijuana Use and False Negative Rates, by Gender, among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	116
6.59. Estimated True Prevalence of Past Year Marijuana Use and False Negative Rates, by Race/Ethnicity, among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	116
6.60. Estimated True Prevalence of Past Year Marijuana Use and False Negative Rates, by Age Group, among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	116
6.61. Estimated True Prevalence of Past Year Marijuana Use and False Negative Rates, by the U.S. Born Indicator, among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	116
6.62. Estimated True Prevalence of Past Year Marijuana Use and False Negative Rates, by Metropolitan Status, among Persons Aged 12 or Older: 2006 NSDUH Reliability Study .....	116
7.1. Impact of the Administration of Both Interviews by the Same Field Interviewer (FI) or by Different FIs on the Reliability of Selected Measures: 2006 NSDUH Reliability Study .....	118
8.1. Impact of the Time between Interviews on the Reliability of Selected Measures: 2006 NSDUH Reliability Study.....	120
C.1. Weighted Distribution of the Responses to the Follow-Up Questions across All Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study .....	139
C.2. Weighted Distribution of Responses to the Follow-Up Questions among Same Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study .....	143
C.3. Weighted Distribution of Responses to the Follow-Up Questions among Different Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study .....	147

# Executive Summary

Information on data quality is an important output of major Federal surveys because survey data often are used to influence policy decisions. A range of types of error may occur in surveys related to problems with the respondent's understanding of the questions or the effects of the interviewer. These problems may bring about a disparity between the survey response and a true value. Reinterviewing survey respondents in studies of survey response reliability or consistency provides a direct measure of such response variance.

This report presents findings from a reinterview study of respondents to the National Survey on Drug Use and Health (NSDUH). The NSDUH Reliability Study responds to a 2006 directive by the Federal Government's Office of Management and Budget to Federal agencies that conduct statistical surveys to evaluate the quality of such surveys and to document the findings of the evaluation to assist users in their interpretation and uses of findings from the surveys.

## Study Objectives

The Reliability Study had two main objectives:

- Provide a comprehensive assessment of the reliability of NSDUH data.
- Identify questions that did not score high on reliability measures and are candidates for further testing and improvement.

Related objectives of the study were to examine the extent to which the reliability of measures varied among segments of the population and to investigate the effect of the interviewer on the reliability of measures.

## Study Procedures

The Reliability Study was embedded within the 2006 NSDUH main study. A subsample of the main study sample of 67,802 was selected such that data from the initial interview were used for both the main study and the Reliability Study. As for the main study, the respondent universe included the civilian, noninstitutionalized population aged 12 or older. However, for the Reliability Study, the respondent universe excluded residents of Alaska and Hawaii, residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories), and persons who did not speak English. To preserve the results and response rates of the main study, neither the field interviewers (FIs) nor respondents were informed ahead of time about the selection for the second interview. Recruitment scripts for the second interview were added to the end of the first interview and administered to the 3,516 eligible respondents selected for reinterview. Second interviews were obtained from 3,136 respondents, for an 85.6 percent reinterview weighted response rate. Although an incentive of \$30 was offered for the first interview, \$50 was offered for participation in the second interview.

The second interviews were conducted 5 to 15 days following the initial interview in order to minimize the chance that respondents would recall their responses to the first interview and to minimize the chance of responses changing for valid reasons. For the Reliability Study, the same questionnaire was used for the second interview as for the first interview, which was conducted as part of the main study. A set of follow-up questions was added to the end of the second interview about respondent use of tobacco, alcohol, and other drugs between the two interviews. A substudy was conducted within the Reliability Study—the same versus different interviewer substudy—to examine the potential impact that an FI might have on reliability.

## **Approaches to Analysis**

To examine the *reliability* of survey responses between the initial and second interviews, as measured by their consistency, the data were edited only lightly to make them suitable for analysis. Analyses compared the following: (a) major items, (b) selected composite variables, (c) select-all variables coded into dichotomous variables, and (d) substance abuse variables that were defined separately from substance dependence variables (in contrast to the abuse and dependence variables in the main study).

Analyses examined the extent of agreement between the two interviews in terms of exact agreement and for selected variables in terms of a more relaxed agreement criterion. Three primary measures of assessing agreement were used: (a) percent reporting consistently, (b) Cohen's kappa ( $\kappa$ ), and (c) the index of inconsistency. All were weighted to provide population estimates.

Models also were fitted to the data to investigate the patterns of nonresponse at the time of the second interview and the relationship between the age at first use and the time since first use. Another model attempted to estimate the true prevalence of past year marijuana use by making certain assumptions about false positives and false negatives among subgroups.

## **Findings**

Comparisons of responses to the initial and second interviews were conducted for selected measures, including substance use, perceived risk and availability of substances, substance dependence or abuse, mental health treatment and mental health problems, health care, health conditions, and demographic characteristics. Using measures and interpretations described in this report, the following were found:

- Responses for substance use in the lifetime had almost perfect reliability, and responses for substance use in the past year showed substantial agreement. Age at first use of specific substances showed mostly moderate reliability, but findings for which substance was used first were less consistent, with some being of only fair reliability. The reliability of responses to age at last use was generally fair.
- Items measuring respondents' perceptions of the risk and availability of substances were usually of moderate reliability.

- Indicators of substance dependence and abuse were of moderate to substantial reliability and were higher for alcohol than for other drugs. Other variables for dependence and abuse were of comparable reliability, as were substance use treatment items.
- Measures of health conditions and for health care usually exhibited substantial agreement.
- Measures of mental health treatment and mental health problems, such as serious psychological distress and depression, were often of only moderate reliability or less. Lower reliability items included location of treatment, source of payment, and health care coverage. Items making up the depression variable were not consistently reliable, although the lifetime depression measure showed substantial reliability.
- Demographic variables showed substantial or better reliability for most measures. Exceptions were having lived in the United States less than 1 year and number of days of school missed.
- Twelve models were fitted to the data to examine whether the ability of respondents to recall age at first use of various substances declines as more years have passed since their first use. Most of the models showed that the consistency of responses for the two interviews declined as the number of years since first use increased.
- Modeling of the prevalence of past year marijuana use suggested that reporting error was greater among the following: (a) blacks and racial/ethnic groups other than whites and Hispanics; (b) youths aged 12 to 17 and adults aged 26 or older compared with young adults aged 18 to 25; (c) those not U.S. born compared with those born in the United States; and (d) persons living in metropolitan areas compared with those living in nonmetropolitan areas.
- Comparisons of the consistency of responses among those who were interviewed by the same versus different FIs at the time of the two interviews showed no significant effect of the interviewer on the reliability of survey responses.
- The consistency of responses among those whose first and second interviews were less than 9 days apart was similar to the consistency of responses among those whose interviews were 9 or more days apart.
- Analyses showed that questions about factual personal events or characteristics were more reliable than questions that asked for a respondent's personal opinion or intentions or questions that addressed issues that carried a social stigma.





# 1. Introduction

Information on data quality is an important standard output of major Federal surveys. Because survey data often are used to influence major policy decisions, users of these data would benefit if data quality were assessed and reported upon in a regular and timely fashion. Most reports from surveys provide data on response rates and sampling error, but they provide little data on measurement error. Response rates, which can indicate potential sources of bias in data, cannot truly measure data accuracy. Sampling error, which is the error caused by selecting a sample instead of collecting data from the entire population, provides some information on the accuracy of estimates. Measurement error refers to a range of types of error related to the disparity between the survey response and a true value. Measurement error may be caused by problems with the survey questions, the respondent's understanding of the questions, or the effects of the interviewer. Reinterviewing respondents in studies of survey response reliability or consistency provides a direct measure of error due to response variance. The capability of the survey to provide accurate data, and consequent population estimates, can be examined by assessing its reliability. Reliability is of particular concern when respondents are asked questions on sensitive topics.

The Office of Applied Studies (OAS) of the Substance Abuse and Mental Health Services Administration (SAMHSA), in its continuing effort to assess and improve the quality of data collected in the National Survey on Drug Use and Health (NSDUH), conducted a study of the reliability of the NSDUH questionnaire. This study was performed in conjunction with the 2006 NSDUH and is referred to in this report as the Reliability Study. The study's main purposes were to (1) provide a comprehensive assessment of the reliability of NSDUH data, and (2) identify questions of lower reliability, thus indicating their potential need for further testing and improvement. Secondary objectives included identifying respondent groups with differing levels of response consistency on a subset of measures and assessing the effect of field interviewers on reliability. This information should help data users better plan their analyses and help them interpret both past and future published findings from the data. The information should also assist in informing questionnaire decisions for the planned NSDUH redesign.

In 2006, the Federal Government's Office of Management and Budget (OMB) issued a directive to Federal agencies that conduct statistical surveys to evaluate the quality of such surveys and publicly document the findings of the evaluation to assist users in their interpretation and uses of findings from the surveys. The implementation of the Reliability Study and the production of this report are intended to serve that purpose for NSDUH.

## 1.1. Summary of NSDUH

NSDUH is the primary source of statistical information on the use of illegal drugs by the U.S. population. Conducted by the Federal Government since 1971, the survey collects data by administering questionnaires to a representative sample of the population through face-to-face interviews at the respondent's place of residence. The survey is sponsored by SAMHSA, which is part of the U.S. Department of Health and Human Services (HHS), and is planned and managed by SAMHSA's OAS. Data collection is conducted under contract with RTI

International, Research Triangle Park, North Carolina.<sup>1</sup> This section briefly describes the survey methodology; a more complete description is provided in Appendix A.

NSDUH collects information from residents of households and noninstitutional group quarters (e.g., shelters, rooming houses, dormitories) and from civilians living on military bases. The survey excludes homeless persons who do not use shelters, military personnel on active duty, and residents of institutional group quarters, such as jails and hospitals.

Since 1999, the NSDUH interview has been carried out using computer-assisted interviewing (CAI). Most of the questions are administered with audio computer-assisted self-interviewing (ACASI). ACASI is designed to provide the respondent with a highly private and confidential means of responding to questions to increase the level of honest reporting of illicit drug use and other sensitive behaviors. Less sensitive items are administered by interviewers using computer-assisted personal interviewing (CAPI).

The 2006 NSDUH employed a State-based design with an independent, multistage area probability sample within each State and the District of Columbia. The eight States with the largest population (which together account for 48 percent of the total U.S. population aged 12 or older) were designated as large sample States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas), with target sample sizes of 3,600 in each. For the remaining 42 States and the District of Columbia, the target sample size was 900. This approach ensured there was sufficient sample in every State to support small area estimation (SAE) while at the same time maintaining efficiency for national estimates. The design oversampled youths and young adults, so that each State's sample was approximately equally distributed among three age groups: 12 to 17 years, 18 to 25 years, and 26 years or older.

Nationally, 137,057 addresses were screened for the 2006 survey, and 67,802 completed interviews were obtained. The survey was conducted from January through December 2006. Weighted response rates for household screening and for interviewing were 90.6 and 74.2 percent, respectively.

## **1.2. Study Objectives**

The Reliability Study, which began in April 2006 and ran through December, involved a subset of the regular NSDUH respondents for that year. The subset was designed to be similar to the main sample with regard to age and gender, but it excluded certain individuals because of cost and other considerations. Selected respondents received a second administration of the NSDUH interview 5 to 15 days after their participation in the main study. The term "reliability" in this report measures the extent to which respondents answered consistently when the same questions were presented on two occasions separated by the specified time period (i.e., between the initial interview, T1, and the second interview, also referred to as the "reinterview," T2).

High reliability is a necessary condition that must be met for data to be considered valid. If a question or set of questions has low reliability, then it is reasonable to infer that one (or both) of the responses is not valid or truthful among a significant proportion of the sample. In addition, if the overall reliability of a measure is found to be good, but its reliability among a specific

---

<sup>1</sup> RTI International is a trade name of Research Triangle Institute.

subpopulation was poor, then any difference in prevalence between the specific subpopulation and others might simply be an artifact resulting from measurement error. The plan was for this study to identify which NSDUH measures are most at risk of producing unreliable estimates in order to enhance understanding of the data, and so that later efforts might be directed toward improving any faulty measures.

There also was a concern that the interviewer may have an effect on the reliability of measures. Thus, a same versus different interviewer substudy was embedded to assess the interviewer effect on measures of reliability. In a random one third of the cases, a different interviewer from the one who completed the initial interview was assigned to complete the reinterview. In the other cases, the same interviewer completed both the initial interview and the reinterview.

### **1.3. Content and Organization of This Report**

This Reliability Study report is divided into nine chapters and four appendices. Chapter 2 offers a summary of other reinterview studies to provide a basis for comparison with this study. Chapter 3 describes the methodology and data collection procedures used for this reinterview study, including the sample design, instrument design, and an explanation of the data prepared for the analyses. Information about a study pretest and the same versus different interviewer substudy also is provided. Chapter 4 details the various statistics and statistical models used to analyze and compare the initial interview responses with the reinterview responses. This includes weighted percent reporting consistently, Cohen's kappa ( $\kappa$ ) statistic, the index of inconsistency (*IOI*), and three types of models, including response propensity models, logit models for age-at-first use and duration of recall, and modeling under the Hui-Walter paradigm for dichotomous outcomes. Chapter 5 describes the survey sample, response rates, and timing of interview data. Chapter 6 provides a general discussion of the analysis tables, including what information is contained in the tables and how they are organized; this chapter also presents the analytical results with an associated narrative to point out the overall findings and any significant findings. Chapter 7 presents findings of interest for the same versus different interviewer substudy. Chapter 8 discusses the reliability findings as a function of the time between the initial interview (T1) and the reinterview (T2). The report ends with Chapter 9 providing recommendations for future reliability studies.

Appendices are included for the reader who is interested in more of the technical details of the study. Appendix A offers a detailed discussion of NSDUH. Appendix B provides a discussion of the kappa statistics, a description of the development of the computation of standard errors for kappa when the data are from a complex survey design, a description of the *IOI* statistic and its variance computation, and a discussion of the rules employed for suppressing the display of the kappa values in the analysis tables. Appendix C presents the results of the T2 follow-up questions that asked respondents how much they had remembered and the extent to which their T1 responses affected their T2 responses. Also, information on the contributors to this report is provided in Appendix D.



## 2. Prior Reinterview Studies and Measures of Reliability

Reinterview studies are conducted for a variety of purposes, including evaluation of interviewer performance, the estimation of bias, evaluation of the questionnaire or data collection instrument, and the estimation of response variance. Forsman and Schreiner (1991, pp. 280-288) pointed out that the design of the reinterview study should depend on the purpose. A variety of reinterview studies are conducted each year in association with data collection for the National Survey on Drug Use and Health (NSDUH). As an example, NSDUH field staff routinely contact a sample of respondents to verify that an actual interview was conducted and to gather limited data about the interviewing environment. These very brief verification interviews are essential for monitoring interviewer performance and for identifying potentially fraudulent interviewing. To evaluate bias, a more accurate, and more expensive, follow-up interview may be required (e.g., a follow-up by clinical experts may be needed to confirm a physical health or a mental health diagnosis). To estimate response variance, the conditions of the original survey should be preserved to the extent possible. Adjustments for time-related questions or elimination of these items from the analysis may be required. Some time is normally allowed between the initial interview and the reinterview so that any nonsystematic errors in the responses can be assumed to be independent of each other. To capture the results of the interviewing and response processes only, very limited editing of initial responses should be employed.

In the Office of Management and Budget (2006) standards and guidelines for statistical surveys, Standard 3.5 directs agencies to "evaluate the quality of the data and make the evaluation public (through technical notes and documentation included in reports of results or through a separate report) to allow users to interpret results of analyses, and to help designers of recurring surveys focus improvement efforts" (pp. ii and 19). Additional "best practice" guidelines specify that both sampling and nonsampling errors should be addressed. Standard 7.3 addresses survey documentation. Guideline 7.3.3 specifies periodic evaluation reports for recurring surveys that itemize all sources of identified errors. The standards document refers to Federal Committee on Statistical Methodology (FCSM, 2001) Working Paper 31 for additional details.

Working Paper 31 discusses a variety of reinterview studies. The section on the simple response variance (*SRV*) reinterview is most applicable to this NSDUH Reliability Study. In an *SRV* reinterview study, the initial interview is replicated to the extent possible, and no reconciliation of results is done. Some common statistics for *SRV* reinterviews include the gross difference rate (*GDR*), the *SRV*, and the index of inconsistency (*IOI*). The *GDR* is the average squared difference between the original and reinterview responses (often weighted to the population). When divided by 2, it can provide an estimate of the *SRV*. The *IOI* is derived from the *GDR* as

$$IOI = \frac{GDR}{s_1^2 + s_2^2},$$

where  $s_1^2$  and  $s_2^2$  are the total variances for the original interview and the reinterview, respectively. The *GDR* represents the proportion of the total population variance due to the *SRV*. The FCSM (2001) Working Paper 31 suggested a common interpretation for the *IOI* as follows:

- low inconsistency for an *IOI* less than 0.20;
- moderate inconsistency for an *IOI* between 0.20 and 0.50; and
- high inconsistency for an *IOI* above 0.50.

Both the *GDR* and *IOI* provide users with information on the reliability of survey questions. Identifying questions that have high inconsistency also provides survey designers with information for selectively improving data quality through further exploratory work to overcome any problems contributing to the high inconsistency measures.

Reinterview studies also can focus on agreements (rather than differences) between the initial interview and the reinterview. Much of the published research on agreement measures in reliability studies focuses on agreement among raters (e.g., interviewers, judges, and expert clinicians) and generally applies to categorical data. The general approach can be applied to agreement in respondent reports at two time points as well as to raters. Simple agreement rates tend to be high for very rare or very prevalent attributes because the probability of agreement from chance alone is high under these circumstances. Cohen's (1960, 1968) kappa ( $\kappa$ ) is an attempt to adjust for chance agreement and is defined as

$$\kappa = \frac{P_o - P_c}{1 - P_c},$$

where  $P_o$  is the proportion of cases observed to be in agreement and  $P_c$  is the proportion that would be expected to agree because of chance alone, generally computed under the assumption of independence of the marginal distributions. Although kappa is a serious improvement over the simple agreement rate, its values still must be interpreted with caution for estimates of proportions near 0.00 or 1.00. Fleiss and Cohen (1973) and Fleiss (1975) pointed out that kappa is analogous to the intraclass correlation coefficient, which is defined for both categorical and continuous data. Landis and Koch (1977) recommended the use of random effects models to test hypotheses about the equivalence of intraclass correlations (or kappas) across subpopulations. They also provided a finer verbal interpretation of kappa statistics (*ibid.*, p. 165):

- poor agreement for kappas less than 0.00;
- slight agreement for kappas of 0.00 to 0.20;
- fair agreement for kappas of 0.21 to 0.40;
- moderate agreement for kappas of 0.41 to 0.60;
- substantial agreement for kappas of 0.61 to 0.80; and
- almost perfect agreement for kappas of 0.81 to 1.00.

More stringent descriptors of levels of agreement are provided by Shrout (1998) as follows:

- virtually no agreement for kappas less than 0.11;
- slight agreement for kappas of 0.11 to 0.40;
- fair agreement for kappas of 0.41 to 0.60;
- moderate agreement for kappas of 0.61 to 0.80; and
- substantial agreement for kappas of 0.81 to 1.00.

Kappa is approximately the complement of the *IOI* for nominal data,<sup>2</sup> so the "almost perfect" (Landis & Koch, 1977) or the "substantial" (Shrout, 1998) levels of agreement for kappa correspond to the low inconsistency level of the *IOI* suggested by the FCSM (2001) Working Paper 31.

This chapter focuses on reinterview surveys conducted to estimate response variance at the item level. Although the analysis in this report is presented primarily in terms of agreement rates and Cohen's kappa, other national surveys discussed below often report their results in terms of *GDRs* and the *IOI*. Some examples and selected results from other major surveys are presented.

## 2.1. U.S. Decennial Census

The Census Bureau's post-enumeration surveys since 1950 have had a component, the P sample, focused on undercoverage and another component, the E sample, based on reinterviewing persons previously covered in the decennial census (Citro & Cohen, 1985, Chapter 4). The timing of the decennial census and the post-enumeration survey required special questions to get comparable household composition. Some of the results from the reinterview component for 1950 and 1960 are presented by Taeuber and Hansen (1964, p. 11). They used a version of the *IOI* (Pritzker & Hanson, 1962) for categorical variables computed by dividing the *SRV* ( $GDR / 2$ ) by an estimate of the total population variance (average of  $p[1 - p]$  for the two interview occasions). For variables with multiple categories, they used an unweighted average of measures for individual categories. This was meant to approximate the portion of the population variance attributable to reporting error. The reported measures for 1960 ranged from 0.018 for sex (i.e., gender) to 0.256 for educational attainment classes.

More recently, Singer and Ennis (2003) reported on the results of the 2000 reinterview survey and related some of the findings to the 1990 results. Approximately 20,000 preselected households were contacted successfully for a reinterview by telephone or personal interviews. Moreover, all had been asked to complete the long form of the census questionnaire. The initial interview was conducted by mailback procedures in about three fifths of these households; in many cases, the reinterview results reflected a change in mode and other factors. The reinterviews occurred over the period from late June through mid-November. Any questions about the last week were dropped from the reinterview, and individual person data were collected for one randomly selected person only. Both the original and the reinterview questionnaire received only minimal editing before the analysis of inconsistency. *IOI* statistics were developed to reflect lack of agreement. The authors categorized inconsistency as low,

---

<sup>2</sup> For nominal data, the *GDR* is 1 minus the agreement rate. A discussion of the relationship between estimates of the *IOI* and kappa can be found in Section B.3.3 of Appendix B.

moderate, or high following the FCSM (2001) categories discussed earlier. The analysis was able to classify 16 items with low inconsistency, 16 items with high inconsistency, and 26 items with intermediate inconsistency. The low inconsistency items included "sex," "age," "Hispanic origin," "marital status," "school attendance," "language spoken at home," "place of birth," "citizenship," "year of entry into the United States," "veteran status," and "period of military service." High inconsistency items included "language usage," "grandparents as caregivers," "work experience in 1999," and "income."

## **2.2. Other Census Bureau Surveys**

Four large surveys conducted by the Census Bureau have reinterview sampling programs. For each, the reinterview sampling rates are shown in parentheses:

- Current Population Survey, CPS (1 in 30);
- National Crime Survey, NCS (1 in 18);
- Survey on Income and Program Participation, SIPP (1 in 18); and
- Consumer Expenditure Survey, CE (1 in 18).

The Census Bureau typically uses supervisors to conduct reinterviews because the goal is to control the quality of interviewing. The use of the telephone mode for reinterviews is maximized even when the original mode of interview was face-to-face (Forsman & Schreiner, 1991, pp. 290-292).

Sinclair and Gastwirth (1996) tested the methodology developed by Hui and Walter (1980) to estimate false positive and false negative rates for CPS estimates of persons in the labor force. They considered the methodology appropriate for estimating error rates from the unreconciled reinterview sample. The reconciled interview sample was used to estimate the net bias. Their analysis showed that estimates based on the reconciled reinterview sample may be no better than the original estimates in terms of misclassification error.

## **2.3. National Household Education Survey**

For the 1993 National Household Education Survey (NHES:93), reinterviews were conducted with 1,879 computer-assisted telephone interviewing (CATI) respondents (Brick, Rizzo, & Wernimont, 1997). They included 227 interviews with parents of 3rd to 5th graders, 277 interviews with parents of 6th to 12th graders, 493 interviews with 6th to 12th graders themselves, and 882 interviews with parents of younger children. Reinterviews were not conducted before 2 weeks following the initial interview. Interviews were about children's experiences, but the same parent was interviewed at both the initial interview and the reinterview. The survey topics were related to school safety and discipline for older children and school readiness for younger children.

Reliability was assessed using the *GDR*, the *IOI*, and correlation. For questions asked of 6th to 12th graders, the *IOI* ratings were in the moderate range (*IOI* between 20 and 50) for "child had drug education," "drug education/regular courses," "drug education/other activities," "students high at school," and "drug dealers at school." High *IOI* ratings (greater than 50) applied to "drug education/special course," "drug education/assemblies," and "students drunk at school."



Median *IOIs* also were reported for item groups related to "general environment," "drug and alcohol education," "drug and alcohol use," and "incidents of victimization at school." For parent reporting, median *IOIs* were all in the moderate range. For student reporting, median *IOIs* were in the moderate range except for "drug and alcohol education," which was rated as high.

## 2.4. Diagnostic and Validation Studies

The studies discussed here are not strictly reinterviews, but they illustrate the use of the kappa statistic in assessing reliability of measurement processes and in the validation of survey-based data against an alternative measure usually presumed to be a gold standard.

Jordan, Karg, Batts, Epstein, and Wiesen (2008) pointed out that the kappa statistic commonly is used to assess the reliability of two raters for nominal scales. The focus is on the consistency of two raters rather than the consistency of the respondent on two occasions. Kulka et al. (1990, p. D-29) reported on a small experiment to check the reliability of clinicians in rating diagnoses of post-traumatic stress disorder (PTSD) based on a taped clinical interview and obtained high agreement (kappa = 0.933); repeating the entire clinical interview process was not considered feasible given the demands on the subject. Jordan et al. (2008) also pointed out that the kappa statistic often is reported in studies that validate a less rigorous diagnosis against a presumed true diagnosis. They reported on the validation of the computer-assisted interviewing (CAI) used in NSDUH against the Structured Clinical Interview for DSM-IV (SCID)<sup>3</sup> when administered by trained clinicians to a recruited sample of 288 respondents from the community and from outpatient substance abuse programs. For 14 selected dependence and abuse measures, kappas ranged from 0.47 to 0.87.

In a study validating reported substance use against biological indicators of substance use, Harrison, Martin, Enev, and Harrington (2007) reported kappas for self-reports of past 30 day substance use against urinalysis. Reported kappas for core questions and urinalysis include 0.643 for any tobacco, 0.517 for marijuana, and 0.243 for cocaine. The study also included a second opportunity to report past substance use after a special honesty appeal providing a reinterview within the same setting with different motivational prompts. Kappas for use in the past 30 days were higher for those who received the appeal than for those who did not.

---

<sup>3</sup> "DSM-IV" refers to the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (American Psychiatric Association [APA], 1994).



## 3. Methods and Procedures

The primary goal of the Reliability Study for the 2006 National Survey on Drug Use and Health (NSDUH) was to provide estimates of response variance (i.e., reliability) of the main NSDUH questions and composite variables. To achieve this goal, the survey instrument was administered to a sample of respondents on two occasions. The second interview (T2) was administered 5 to 15 days after the initial interview (T1). The reliability of the instrument was measured by the degree of agreement between the survey responses from the two administrations.

### 3.1. Overview of Reliability Study Procedures

The Reliability Study sample was embedded within the 2006 NSDUH main study sample. A subsample of the main study sample was selected such that data from the initial interview were used for both the main study and the Reliability Study. To preserve the results and response rates of the main study, the Reliability Study was double-blinded. Neither field interviewers (FIs) nor respondents were aware beforehand which respondents would be selected for a second interview. Once the main study interview was complete, the computer-assisted interviewing (CAI) program applied an algorithm to determine whether the respondent was selected for the Reliability Study. If the respondent was chosen, the CAI program would present the Reliability Study recruitment scripts to the FI. This was the first moment an FI would know that a respondent would be asked to participate in the Reliability Study. In addition, FIs were instructed not to mention the possibility of a second interview at any time during the main study recruitment phase. At the conclusion of the main study interview, selected respondents were simply asked to participate in a second interview "to help us improve our interviewing procedures and how we ask questions." At no time were respondents informed that the second interview would be the same as the first one.

#### 3.1.1 Time between T1 and T2 Administrations

To ensure the validity of the results, the amount of time that elapsed between the first (designated the T1 interview) and second (designated the T2 interview) administrations of the interview was considered. Enough time needed to pass so that respondents could not recall their responses to the initial interview. The upper limit to the window of time also was fixed to minimize the chance of responses changing for valid reasons. Thus, it was determined that reinterviews should be conducted sometime during the period from 5 to 15 days following the initial interview.<sup>4</sup>

#### 3.1.2 Same versus Different Interviewer Substudy

One concern raised during discussions of the Reliability Study design was the potential impact an FI might have on the follow-up, or T2, interview. To allow some assessment of interviewer effect, a substudy was imposed on the Reliability Study sample where one third of

---

<sup>4</sup> A few reinterviews were conducted outside the 5- to 15-day window. These cases were reviewed on a case-by-case basis to determine whether they should be included in the study.

the cases were assigned to have the T2 interview conducted by an interviewer who was different from the interviewer who conducted the initial, or T1, interview. The other two thirds of the cases were designated to have the T2 interview conducted by the same interviewer who conducted the T1 interview. For cases in which a different FI was designated to conduct the T2 interview, the field supervisor (FS) responsible for the data collection operations in the area where the cases were located selected an FI from the pool of FIs available nearby so as to minimize follow-up interview costs.

### **3.1.3 Incentive**

In the NSDUH main study, a \$30 incentive was provided to respondents who completed the interview. An additional \$50 incentive was provided for respondents who completed the T2 interview of the Reliability Study. The increased incentive was deemed necessary for several reasons, including the following:

- the time frame for completion of the T2 interview was short and restrictive (5 to 15 days after the T1 interview);
- the respondent had already completed the initial NSDUH interview (T1) and had been told the amount of time needed to complete the reinterview; and
- no recontact with refusals would be made in an attempt to convert refusals.

For these reasons, a \$30 incentive was considered insufficient for the T2 interview.

## **3.2. Pretest**

A Reliability Study Pretest was performed to guide the design and implementation of the Reliability Study. The pretest was conducted in two phases with 200 total reinterview respondents in quarters 1 and 2 of the 2005 NSDUH. Phase 1 was conducted in January and February 2005, and Phase 2 in April and May 2005. March 2005 was used to assess the Phase 1 experience, analyze the results, and make refinements to the procedures, materials, and instrumentation for the Phase 2 completion. For a description of the pretest, see Piper, Meyer, and Snodgrass (2006).

The primary purpose of the Reliability Study Pretest was to test the field procedures, materials, and instrumentation, as well as respondents' reactions to the reinterview. Thus, the pretest also examined response rates, methods, and feasibility of administration.

### **3.2.1 Pretest Methodology**

Each phase of the Reliability Study Pretest was conducted using segments that were totally separate from the 2005 NSDUH main study segments. A total of 12 retired NSDUH segments were selected in 4 States in each phase of the pretest for a total of 24 segments making up the nonprobability-based sample. The sample was selected in States where cases could be observed easily by staff from the Substance Abuse and Mental Health Services Administration (SAMHSA) or by NSDUH field management staff. As part of the pretest process, 10 FIs were observed by SAMHSA and NSDUH field management staff completing 13 screenings, 7 initial

interviews, and 7 reinterviews. The pretest was conducted by 23 FIs who were specifically identified and selected based on having a history of good performance on NSDUH. The FIs were trained on the Reliability Study Pretest procedures in person during a half-day training session. FIs also completed a home study exercise and refresher training session via telephone prior to the start of Phase 2.

Dwelling units (DUs) were selected randomly within these 24 designated segments. Within sample households that completed the screening interview, the sampling algorithm programmed on the iPAQ (a handheld computer) to select eligible residents within a cooperating household was designed to select 0 or 1 person. If a second person could be chosen in a Reliability Study household, then the second person could learn from the first person about the follow-up interview, and that would compromise the purpose of the experiment.

As in NSDUH's main study, persons eligible for the pretest were civilian, noninstitutionalized persons aged 12 years old or older. However, the pretest did not include main study eligible persons living in group quarters nor persons who could complete the interview only in Spanish. If a respondent completed the T1 interview in Spanish, then the recruitment scripts at the end of the T1 interview did not appear; thus, the respondent was not recruited for a T2 follow-up interview.

A special same versus different interviewer substudy was imposed on the pretest sample to allow some assessment of interviewer effect and to test associated procedures. Two thirds of the cases were randomly assigned to have the same interviewer do both the T1 and T2 interviews. For the remaining one third of the cases, a different interviewer was required to conduct the T2 interview.

Data collection began with the T1 interview conducted exactly as a main study interview was conducted. At the end of the T1 interview, the CAI questionnaire program prompted the FI with wording designed to recruit respondents for a T2 interview. The interviewer made no mention of the follow-up interview prior to this point. The CAI questionnaire program also determined whether the T2 interview was to be conducted by the same FI or by a different FI. At the close of the T1 interview, the respondent received a \$30 cash payment just as with the main study.

Both T1 and T2 interviews were identical to the main study interview except at the very end. The T1 questionnaire ended with the recruitment scripts mentioned above that included mention of a \$50 payment for completing the T2 interview, and the T2 questionnaire ended with additional questions for the respondent and the FI. The \$50 payment for completing the T2 interview was given to the respondent after the completion of the T2 interview.

Because of the small sample and the need to obtain as much information from respondents as possible regarding the pretest process, all pretest interviews in both Phase 1 and Phase 2 were verified via the same verification procedures used on the NSDUH main study. Cases where both T1 and T2 interviews were completed were verified only once following the completion of the T2 interview.

Following each phase of the pretest, telephone debriefings were conducted with the FIs to probe them on their experiences with conducting the Reliability Study interviews and to obtain their opinions and suggestions for what worked well and what improvements in procedures could be made. Input from these FI debriefing sessions was used to augment the pretest analysis findings and to refine materials and procedures for the 2006 Reliability Study.

Phase 1 of the pretest indicated that no changes needed to be made to the data collection materials, instruments, or staff between the two phases. Following Phase 1 of the pretest, it was decided that the full Reliability Study for 2006 should be conducted as an embedded sample design rather than as a separate sample. Thus, the initial Reliability Study interviews would be included with all other main study NSDUH interviews. The Phase 1 pretest sample and procedures were designed for a sample that was totally independent of the main study sample. This meant that FIs conducting Reliability Study pretest interviews knew that every interview completed in the pretest was a Reliability Study case and would require an attempt to conduct a follow-up interview. This would not be true when the Reliability Study sample was embedded with the main study sample.

It was decided to keep the Phase 2 pretest sample independent of the main study's quarter 2 sample by using 12 retired 2004 NSDUH segments as originally planned rather than trying to simulate an embedded design for Phase 2 of the pretest, which would have been time-consuming, costly, and possibly delayed the start of the full Reliability Study in quarter 1 of 2006. However, to simulate that not every case selected would be a Reliability Study case, there were some sample dwelling units (SDUs) in Phase 2 that were not designated as Reliability Study cases and thus did not require follow-up interviews. This increased the number of DUs selected for Phase 2 to 34 DUs per segment for a total of 408 SDUs.

Phase 2 data collection was originally scheduled to run from May 1, 2005, through June 30, 2005. However, the Phase 2 data collection period was changed to start on April 1, 2005, and run through May 31, 2005, because Phase 1 went very well and there were only minor changes needed for Phase 2. This change allowed for more time to analyze the pretest data and develop the 2006 Reliability Study design.

### **3.2.2 Implications of the Pretest for the Full Study Implementation**

The overall interview response rate for the Reliability Study Pretest increased slightly from Phase 1 to Phase 2. The Phase 1 response rate for the T1 interview (79.4 percent) was somewhat lower than the design parameter of 82 percent, while the T2 interview response rate of 92.9 percent was slightly greater than the design parameter of 92 percent. The Phase 2 T1 interview response rate of 83.2 percent was somewhat higher than the design parameter of 82 percent, whereas the T2 interview response rate of 90.8 percent was somewhat lower than the design parameter of 92 percent. Table 3.1 shows the response rate information for Phases 1 and 2 of the Reliability Study Pretest.

**Table 3.1. Pretest Response Rates, by Phase: 2006 NSDUH Reliability Study Pretest**

<b>Phase</b>	<b>Interview Response Rate (T1)</b>	<b>Reinterview Response Rate (T2)</b>	<b>T2 Refusals</b>	<b>Unable to Contact for T2</b>
Phase 1	79.4%	92.9%	5.4%	1.8%
Phase 2	83.2%	90.8%	5.0%	3.4%

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study Pretest (*n* = 200).

Based on experiences encountered and information gathered during the Reliability Study Pretest, the following conclusions were made for the full study implementation. Other aspects of the pretest remained the same and were implemented in the Reliability Study in 2006.

- The sample for the full Reliability Study would be embedded in the main study sample in 2006. This meant that not every SDU would be designated for the Reliability Study. An embedded design was selected mainly to reduce costs, but also to reduce potential interviewer effects because interviewers would be blinded to which cases would be Reliability Study cases. FIs reported that the addition of nonreliability cases to their assignment in Phase 2 did not create any obstacles to following procedures as long as they read the screens verbatim and used the interviewing tools provided to them.
- Minor improvements would be made to the instrumentation and materials to address the issue that it was more difficult to complete reinterviews in the different interviewer condition because of scheduling logistics. For organization purposes, the reinterview payment receipt would be printed on blue paper so there would be a visual difference between the T1 and T2 payment receipts.
- For the different FI condition cases, respondents did not voice any concerns about providing a phone number or having a different FI return to complete the reinterview, even if it was an interview with their child. However, it was decided that the first FI would not give out any information to the respondent as to the gender or name of the second FI who would return to complete the reinterview. The initial FI would leave a blank appointment card with the respondent, to be completed by the respondent when the second FI called to schedule an appointment for the reinterview.
- Respondent reaction to the reinterview was positive. Respondents rarely had questions, but when they did, FIs were able to successfully address the questions and concerns based on answers provided during training and in the handbook. For the 2006 Reliability Study training, additional guidance would be provided on answering respondents' questions based on pretest experiences.
- A specific protocol would be provided for situations where the parent is available at the start of the initial interview, but then leaves during the interview before the reinterview recruit process (without notifying the FI).

- The \$50 incentive payment seemed to make respondents eager to complete the reinterview. It was felt that offering a smaller incentive would likely result in respondents asking more questions and having to be persuaded to participate in the reinterview, as well as a greater percentage of respondents breaking their appointments. It was decided to keep the \$50 incentive amount for the full Reliability Study.
- Scripts for making T2 interview appointments would be provided to guide FIs when making follow-up contacts with respondents selected for a different FI reinterview condition by telephone.

### 3.3. Sample Design

The Reliability Study sample was embedded within the main study sample. As in the main study, persons eligible for the Reliability Study were the civilian, noninstitutionalized population aged 12 years old or older. Unlike the main study, the respondent universe excluded residents of Alaska and Hawaii, residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories), and persons who did not speak English.

A probability sample of 400 (out of 876)<sup>5</sup> State sampling (SS) regions was selected for the Reliability Study. Probability proportional to size (PPS) sampling was used to obtain a nationally representative sample of SS regions. This sampling method has the effect of equalizing sample weights, at least within age groups (Chromy, 1979). Prior to selecting the sample, implicit stratification was achieved by sorting the SS regions by State and then region within State. From this well-ordered frame, a sample of 400 SS regions was selected with probabilities proportionate to a composite size measure and with minimum replacement. A stratum and replicate structure was developed to account for the subsampling design in variance estimation. Using the sort order from the sampling frame, SS regions were geographically grouped to form strata. Each stratum contained two replicates, with replicates being individual SS regions.

In each of the 400 SS regions, two area segments were sampled in each of quarters 2 through 4 of the 2006 NSDUH survey year. Thus, the Reliability Study sample was drawn from 800 area segments per quarter or a total of 2,400 segments.

Each quarter, a sample of DUs in the 800 segments was designated for the study. Because the study was embedded within the 2006 NSDUH main study, SDUs first were selected for the main study, then a random sample of SDUs was assigned to the Reliability Study. The study was limited to households in which only one person was selected. In households with two persons selected, the second person could learn from the first person about the follow-up interview, which could affect the second person's likelihood of participation and compromise the results of the study. Because eligibility for the study was based on the screening result, the screening instrument was programmed to communicate the eligibility status with the CAI instrument. If one person was selected in a household designated for the Reliability Study, the screener

---

<sup>5</sup> Alaska and Hawaii were excluded from the Reliability Study. Thus, the main study's 900 SS regions were reduced to 876.



generated a questionnaire identification number that triggered an initial (T1) Reliability Study interview in the CAI. Otherwise, the household was ineligible for the Reliability Study and the screening instrument generated a questionnaire identification number that triggered only a main study interview. Section 3.4 describes the difference between the T1 and main study questionnaire versions.

Prior to recruiting the respondent for the second interview, an additional Reliability Study eligibility requirement was checked within the CAI instrument. If the questionnaire was administered in Spanish, the respondent was ineligible for the Reliability Study and the instrument skipped over the recruitment questions. This eligibility requirement was adopted because of the additional costs associated with development of instrumentation in Spanish.

Tables 3.2 and 3.3 show that approximately 26,098 selected main study DUs were expected to yield a total of 3,100 completed reinterviews. Using experience from prior NSDUHs and the 2005 NSDUH Reliability Study Pretest (see Section 3.2), it was assumed there would be a 91 percent screening response rate (SRR) among eligible DUs, an 82 percent interview response rate (IRR) for T1 interviews, and a 92 percent IRR for T2 interviews. The expected overall response rate (ORR) was 69 percent. Additionally, it was assumed that 3 percent of respondents would be ineligible for the T2 interview because they completed the first interview in Spanish. Finally, it was assumed that 84 percent of the DUs would be eligible for the main study, and 21 percent of those would result in a single person selection; thus, the expected eligibility rate for Reliability Study DUs was 18 percent. Assuming a design effect of 1.7, the effective sample size for the Reliability Study was expected to be approximately 1,800.

**Table 3.2. Design Parameters at the Dwelling Unit Level: 2006 NSDUH Reliability Study**

<b>Total Sample</b>	<b>Rate</b>	<b>Number</b>
State Sampling (SS) Regions	N/A	400
Segments	N/A	2,400
Selected Lines	N/A	26,098
Expected Eligible Dwelling Units (DUs) <sup>1</sup>	0.18	4,698
Expected Completed Screening Interviews	0.91	4,275

N/A: Not applicable.

<sup>1</sup>Based on prior NSDUH experience, it was expected that 16 percent of the selected DUs would be ineligible (institutional, nonresidential, etc.). Of the eligible DUs, it was expected that 21 percent would result in a single person selection. Thus, the 26,098 DUs were reduced to 4,698 eligible DUs.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 3.3. Design Parameters at the Person Level: 2006 NSDUH Reliability Study**

<b>Total Sample</b>	<b>Overall Rate</b>	<b>Overall Number</b>	<b>Age 12-17 Rate</b>	<b>Age 12-17 Number</b>	<b>Age 18-25 Rate</b>	<b>Age 18-25 Number</b>	<b>Age 26+ Rate</b>	<b>Age 26+ Number</b>
Expected Selected Persons (First Interview)	1.00	4,275	1.00	1,288	1.00	1,351	1.00	1,636
Expected Completed Interviews (First Interview)	0.82	3,488	0.89	1,140	0.84	1,133	0.74	1,214
Expected Selected Persons (Second Interview) <sup>1</sup>	0.97	3,372	0.99	1,123	0.96	1,088	0.96	1,161
Expected Completed Interviews (Second Interview) <sup>2</sup>	0.92	3,100	0.92	1,033	0.95	1,033	0.89	1,033
Effective Sample Size Based on Assumed Design Effect	1.70	1,824	1.70	607.8	1.70	607.8	1.70	607.8

<sup>1</sup> Because the second interview was not conducted in Spanish, respondents who completed the first interview in Spanish were not selected to complete the second interview. The selection rate for the second interview is based on the number of Spanish interviews completed in 2003, as shown in Tables 7.28 and 7.30 of the data collection final report for the 2003 NSDUH (Bowman et al., 2004).

<sup>2</sup> Rates are based on actual experience during Phases 1 and 2 of the 2005 Reliability Study Pretest (Piper et al., 2006).

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

In the implementation, the 26,098 DUs were distributed among quarters 2 through 4 in proportions of 35, 35, and 30 percent, respectively. Each quarter, the Reliability Study sample was allocated to the 800 segments proportionally to the main study sample size (excluding sampled group quarters units).

To allow greater control over the Reliability Study sample, the DUs were drawn from specific main study partitions.<sup>6</sup> In quarter 2, the Reliability Study sample was limited to partition 1 (80/120 of main sample). In quarters 3 and 4, the Reliability Study sample was drawn from partitions 1 (80/120 of main sample), 3 (5/120), and 6 (10/120). In all three quarters, partition 1 was released. In quarters 3 and 4, consideration was given to the Reliability Study when releasing partitions 3 and 6; however, the main study sample was given priority.

As mentioned in Section 3.1.2, the Reliability Study included a same versus different FI substudy. In some cases, the T1 interview and T2 interview were conducted by the same interviewer, and in other cases the T2 interview was conducted by a different FI working in a nearby area. In quarter 2, one third of the SS regions were assigned to the different FI condition, and the remaining regions were assigned to the same FI condition. In quarter 3, a different one third of the regions were assigned to the different FI condition. Finally, in quarter 4, the remaining one third of the regions were assigned to the different FI condition so that each Reliability Study SS region was in the different FI condition for one quarter and in the same FI condition for two quarters.

<sup>6</sup> Sample partitioning is a standard practice implemented in NSDUH to compensate for quarterly variations in response rates and yields. The full quarterly sample (a 100 percent sample plus a 20 percent reserve sample) is partitioned into random subsamples. Each partition of the sample is a valid sample and helps to control the amount of nonresponse without jeopardizing the validity of the study (Morton, Chromy, Hunter, & Martin, 2007-2008).

A new income module was tested during the administration of the 2006 NSDUH. The Reliability Study was included in this test. Approximately 50 percent of the Reliability Study interviews were assigned the standard income module, and 50 percent were assigned the new income module. Each Reliability Study respondent received the same income module at the T1 and T2 interviews.

### **3.4. Instrument Design**

For the 2006 NSDUH Reliability Study, the screening interview and the CAI questionnaire were set up for an embedded study by completely integrating items specific to the Reliability Study within the main study systems.

***Screener.*** The 2006 NSDUH screening application included additional screens to monitor the response rates for Reliability Study interviews, track the progress of conducting the T2 interviews, and record any reasons for refusal that a respondent may have had at any stage of the interview or reinterview process. T2 cases were appropriately tagged so that field staff and analysts could differentiate them from T1 cases.

***Questionnaire.*** The main study CAI questionnaire program was modified to include Reliability Study items for both the T1 and T2 interviews. If a respondent was selected for the Reliability Study, the main study interview became the T1 interview for the Reliability Study. Respondents selected for the Reliability Study were recruited into the study at the end of the T1 interview using specially prepared recruiting scripts that were programmed appropriately into the CAI questionnaire and administered as computer-assisted personal interviewing (CAPI) items. CAPI screens appeared on the laptop, guiding the FIs through the recruitment process. This process included telling the respondent about the additional study, getting parental consent (if the selected respondent was a minor), and scheduling the T2 interview if the same FI was assigned to conduct it. If another FI was designated, the T1 FI asked for the respondent's phone number and let him or her know that the second FI would be in touch to schedule the T2 interview. Respondents were told that the additional interview was being conducted in order "to help us improve our interviewing procedures and how we ask questions."

At the beginning of the T2 interview, FIs were reminded to read the introduction to the CAI from the special study section of their showcard booklet instead of their job aids booklet. After that, the CAI questionnaire remained the same for the main study and for the Reliability Study's T1 and T2 interviews until the end of the interview. A section was included at the end of the demographics section of the T2 interview to get further feedback from respondents who participated in both Reliability Study interviews. At this point, the FI was directed to turn the laptop back over to the respondent so that he or she could complete the follow-up items using the audio computer-assisted self-interviewing (ACASI) methods again. The follow-up questions asked of the respondents are listed in Table 3.4.

At the end of an interview, the screens prompting the FIs to pay the respondent and complete the quality control form and the incentive receipt form included wording tailored to the type of interview just completed—a main study interview, a T1 interview, or a T2 interview.

### **3.5. Administration of the Questionnaires**

The Reliability Study took place in quarters 2 through 4 of the 2006 survey in the 48 contiguous States plus Washington, DC. Alaska and Hawaii were excluded from the target population because of potential logistical and cost complications associated with the same versus different interviewer design component.

All veteran FIs who worked in the 48 contiguous States were trained on the Reliability Study procedures during a little more than a half day of training at the end of the regularly scheduled 2006 veteran FI training session. New FIs were provided with similar training at the end of the regularly scheduled new-to-project training sessions in 2006. Because the average expected number of Reliability Study interviews completed each quarter by an FI was small, FIs also completed a computerized Reliability Study refresher training course prior to the start of each quarter of data collection to refresh them on the proper procedures and protocol. FIs were provided a separate Reliability Study handbook that contained the detailed study procedures and protocol for reference and review. To maintain existing lines of communication, FIs with Reliability Study issues or questions reported to their regularly assigned FS.

Reliability Study T1 interviews were administered using the same equipment and questionnaire as the 2006 NSDUH main study except that reinterview (T2) recruitment scripts were added at the end of the interview and FIs answered a few additional debriefing questions on the reinterview recruitment process. The T2 interview materials were similar to the main study materials with slight modifications due to the Reliability Study sample size and the \$50 incentive being provided for completion of the T2 interview.

The T2 interview had to be completed within a specific window of time: 5 to 15 days after the completion of the initial interview. Near the end of each quarter, when more than 5 but fewer than 15 days remained in the field period, the last day of the T2 interview time window was prefilled with the date of the last day of the field period; thus, the window of time was shorter than the average 10-day time frame. If the T1 interview was completed when there were fewer than 5 days remaining in the field period, the recruitment scripts did not appear during the T1 interview because the T2 interview could not be completed 5 to 15 days after completing the initial interview.

At the end of the T1 interview, the FI read a CAPI script describing the reinterview process to respondents. If the respondent was 12 to 17 years old, the FI read a CAPI parental consent script to the parent or guardian first. Both the respondent and FI were unaware that the respondent would be asked to complete another interview until these scripts were activated. One of the most important protocols on the Reliability Study was that the respondent would not know about the possibility of a second interview until the recruitment screens appeared at the end of the T1 interview. FIs were to never mention this possibility of a second interview and subsequent interview payment of \$50 while trying to obtain respondent participation for the screening or initial interview. FIs also answered respondent questions about the content of the T2 interview in a standard way to limit the introduction of interviewer bias.

**Table 3.4. Follow-Up Audio Computer-Assisted Self-Interviewing (ACASI) Questions in the T2 Interview: 2006 NSDUH Reliability Study**

Question Text	Response Categories
During the time between the first and second interviews, did you think about your use or nonuse of tobacco, alcohol, and other drugs more than usual, about the same as usual, or less than usual?	More Than Usual About the Same as Usual Less Than Usual DK/REF
How many of the questions in this interview do you think were the <b>same</b> as the questions in the first interview?	All of Them Most of Them Some of Them None of Them DK/REF
Please think about the questions on your use or nonuse of tobacco. How many of your answers to the tobacco questions do you remember from the <b>first</b> interview?	All of Them Most of Them Some of Them None of Them DK/REF
How many of your answers to the tobacco questions in the second interview were the same as your answers to the tobacco questions in the first interview?	All of Your Answers Were the Same Most of Your Answers Were the Same Some of Your Answers Were the Same None of Your Answers Were the Same DK/REF
Please think about the questions on your use or nonuse of alcohol. How many of your answers to the alcohol questions do you remember from the <b>first</b> interview?	All of Them Most of Them Some of Them None of Them DK/REF
How many of your answers to the alcohol questions in the second interview were the same as your answers to the alcohol questions in the first interview?	All of Your Answers Were the Same Most of Your Answers Were the Same Some of Your Answers Were the Same None of Your Answers Were the Same DK/REF
Please think about the questions on your use or nonuse of marijuana. How many of your answers to the marijuana questions do you remember from the <b>first</b> interview?	All of Them Most of Them Some of Them None of Them DK/REF
How many of your answers to the marijuana questions in the second interview were the same as your answers to the marijuana questions in the first interview?	All of Your Answers Were the Same Most of Your Answers Were the Same Some of Your Answers Were the Same None of Your Answers Were the Same DK/REF
Now think about all of the questions in both interviews. Overall, would you say that your answers were more accurate in the first interview, more accurate in the second interview, or about as accurate each time?	My Answers Were More Accurate in the First Interview My Answers Were More Accurate in the Second Interview My Answers Were Just as Accurate in Each Interview DK/REF

DK = don't know; REF = refused.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

At the end of the T1 interview, the FI followed step-by-step instructions included on the recruitment screens to activate the T2 interview case within the iPAQ screening device. Case activation established and automatically set the specific window of time for the T2 interview to be completed and allowed the FI to enter record of call data into the iPAQ for the T2 interview case.

In the same FI condition, the interviewer recruited the respondent for the T2 interview immediately following completion of the T1 interview and made an appointment for 5 to 15 days later; moreover, the respondent received a reinterview appointment card. In the different FI condition, the interviewer recruited the respondent for the T2 interview immediately following completion of the T1 interview and collected his or her phone number so the second FI could call to set up an appointment for the T2 interview. The FI entered this contact information into the password-protected iPAQ record of calls for the T2 interview and left a blank reinterview appointment card with the respondent to complete when the second FI called to set up the appointment. The FI calling to set up the appointment was provided contact scripts to use as a guide through the telephone call with the respondent.

### **3.6. Refusals, Consent, and Incentives**

If the respondent refused to complete the T2 interview at recruitment or at the time of the T2 interview, the FI addressed any questions the respondent had at that time, but did not return to the household at a future time to try to convert the refusal. If the FI was unable to gain respondent participation for the T2 interview, the case was finalized as a refusal. Refusal letters, as used on the NSDUH main study, were not used for T2 interview refusal cases. Similarly, if the FI was unable to contact a T2 interview case, an unable to contact form letter was not sent to these Reliability Study reinterview cases.

At the beginning of the T2 interview, the FI read the reinterview introduction and informed consent statement to the respondent, which instructed the FI to provide the respondent a copy of the special study description. Verbal parental or guardian consent was obtained for any 12 to 17 year old selected for the T2 interview. This text was included in the introduction and informed consent for reinterviewing respondents aged 12 to 17.

After the T2 interview was completed, each respondent received a \$50 cash payment and an FI-signed reinterview payment receipt. For verification purposes, respondents were asked to complete a reinterview quality control form that requested their address and telephone number so that RTI (the data collection contractor) could verify that the interviewer did his or her job appropriately. Last, respondents were asked to complete the follow-up questions shown in Table 3.4 via ACASI. These questions obtained information on topics ranging from whether respondents had thought about their use or nonuse of tobacco, alcohol, and other drugs more or less than usual to whether they remembered their initial interview answers and whether their reinterview answers were the same. The FI then went on to finish the end of interview procedures just as with the initial interview cases. The FI answered the standard NSDUH interview FI debriefing questions and entered a code 70 (Interview Complete) for the T2 interview case.

### **3.7. Response Rates and Verification**

The Reliability Study resulted in a 74.9 percent T1 interview weighted response rate and an 85.6 percent T2 interview weighted response rate. The T1 interview weighted response rates for the same versus different FI substudy were very similar—74.7 percent for the same FI condition and 75.2 percent for the different FI condition. The T2 interview weighted response rates for the same versus different FI substudy were more separated—89.3 percent for the same FI condition and 78.3 percent for the different FI condition. Thus, having the same interviewer do both interviews did increase the response rate for the follow-up interview. However, this effect should not be attributed solely to respondent preferences. The different FI condition entailed additional contact procedures (see Section 3.1.3), which may have reduced the likelihood of T2 participation. For more detailed information on Reliability Study response rates, see Section 5.2.

As with the 2006 NSDUH main study, verification was completed on 5 percent of the noninterview screening respondents. A telephone interviewer called the screening respondent to verify that the screening occurred and whether it was conducted following proper project procedures.

In quarter 2 of 2006, due to the small number of Reliability Study cases each FI would complete and to ensure that Reliability Study procedures were being properly implemented, verification was completed on all respondents who completed the initial interview and/or reinterview and completed the quality control form. For the remaining quarters of 2006, verification was completed on 50 percent of the respondents who completed the initial interview and/or reinterview and completed the quality control form. Respondents who completed both interviews only received one completed verification call to verify the reinterview.

If a respondent refused to complete the T2 interview or was unable to be contacted for the T2 interview, that respondent was asked a few follow-up questions at the end of the verification process for the initial interview. There were separate follow-up questions for respondents who refused during the recruiting process after completing the T1 interview and for respondents who initially agreed to participate during the recruiting process and then refused at a later time.

A verification letter was mailed when a telephone number was not provided on the quality control form for a T1 or T2 interview case selected for verification. The respondents who completed the T2 interview received a letter similar to the NSDUH main study verification letter except that it focused on the T2 interview protocol. Respondents who completed only the T1 interview received the main study verification letter and nothing else.

### **3.8. Analysis Data**

The dataset for the analyses was constructed by matching T1 and T2 responses from the same individual, thereby creating a single record for each subject who had responded to both interviews. The resulting dataset consisted of 3,136 usable records. For calculating reliability measures on any given variable, records with missing values or legitimate skips for the variable at either interview were omitted.

### 3.8.1 Types of Variables Examined

The normal NSDUH data processing operations include extensive editing and imputation procedures applied to the raw data to correct errors and fill in missing responses. However, to assess the reliability of questionnaire items, using the raw data responses in the analysis is more useful than data that have undergone editing and imputation, which may blur some of the differences between T1 and T2 data and mask their true reliability.

**Main Variables.** Using strictly raw variables was not practical for a meaningful reliability analysis. Thus, the raw data were lightly edited to make them suitable for analysis. This light editing included assigning numerical codes to certain classes of responses, collapsing all missing and unknown categories to a SAS<sup>®</sup> missing code, and performing similar types of consistency edits.

Another type of data editing was the combination of information collected from a series of follow-up, probing, and appeal-type questions used when a refusal or "don't know" response to the original question was obtained. That is, rather than looking at the reliability of the original and repeated questions separately, the series of questions were used to determine a final outcome and that final outcome was analyzed. This was common for lifetime substance use variables where a respondent may refuse to answer the initial question about ever using a particular substance, but does provide a definitive response in one of the follow-up questions. Thus, the analyses in this report were conducted on a minimally processed data file.

For the purpose of analyzing reliability by domains (e.g., education, race/ethnicity, or age groups), the T1-imputed value of the domain variable was used when necessary. Because the T1 interview also was included with the full set of main study interviews, imputation of the demographic domain variables for T1 respondents was done as part of the main study imputation process.

In addition, there was an analysis of the reliability of the domain variables. That analysis used the "raw" domain variables. Thus, both the raw domain variables and the imputed domain variables were included on the reliability study data file.

**Composite Variables.** Certain variables in the NSDUH data are derived from combinations of multiple response items. Such variables are referred to as "composite variables." For example, dependence on alcohol is derived from seven criteria, where each of the criteria is based on one or more questionnaire items. Because these composite variables are the result of combinations of multiple other variables, the reliability of the individual variables that make up the composite variables was of interest and is shown in the same analysis table with the composite variables.

**Select-All Variables.** Certain questions in the NSDUH instrument require the respondent to select all the applicable items ("select all that apply"). For example, the respondent is asked to select all race categories to which he or she belongs or all sources of mental health treatment he or she has received in the past year. Each possible choice is either selected or not, so each response option is treated as a separate dichotomous variable. Continuing the example above, a dichotomous variable was created that indicated whether the respondent selected "black." Note



that there is also a multilevel race variable that is created with a "two or more race" category to account for multiple selections. Most "select all that apply" variables are used only to create dichotomous variables.

***Substance Abuse Variables.*** The substance abuse variables in this report were defined simply as meeting at least one of the four criteria for abuse included in the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV) (American Psychiatric Association [APA], 1994). This is different from how abuse is defined in the 2006 NSDUH main study (Office of Applied Studies [OAS], 2007). Persons who meet the DSM-IV criteria for both dependence and abuse are classified by the 2006 NSDUH as having dependence but not abuse. To avoid contamination of the reliability assessment of the abuse classification used in the 2006 NSDUH with the reliability of the dependence classification, abuse for this reliability analysis was defined independently from dependence.

### **3.8.2 Agreement between T1 and T2 Responses**

Agreement between T1 and T2 responses was defined generally as responses that were identical. Also, for certain variables, a more relaxed agreement criterion was applied. For example, a difference of 1 year in the reported age at first use of a substance was allowed under a relaxed criterion of agreement. Certain variables with multilevel responses had a relaxed criterion of agreement if the T1 and T2 responses were just one response category off in either direction. Certain other variables had a relaxed agreement criterion where differences greater than one unit were allowed, such as the reported frequency of use of substances. The analysis tables present variables with exact and approximate or relaxed agreement.

### **3.8.3 Interview Timing Data**

Data on the duration of the entire interview and on the duration of just the ACASI part of the interview for both the T1 and T2 interviews were included in the analysis. Duration data for a few cases were incorrect due to a break-off in the middle of the interview and resumption and completion of the interview a day or more later. Such cases were excluded from the timing analyses.

The number of days separating the two interviews could affect the reliability of the responses (e.g., due to recall). This possibility was examined as well.



## 4. Measures of Agreement and Approaches to Modeling

This chapter describes the three measures of agreement and the approaches to modeling that were used to analyze the Reliability Study data for the 2006 National Survey on Drug Use and Health (NSDUH). Results of these analyses are presented in Chapter 6.

### 4.1. Measures of Agreement

The possible combinations of outcomes for the initial interviews (T1) and the follow-up interviews or reinterviews (T2) are shown in Table 4.1 for an example analysis of a dichotomous variable, "substance use." This simple case is typical of the types of comparisons of NSDUH measures at T1 and T2 that are addressed in this Reliability Study report. Values in the table represent population proportions, that is to say, values that would result were the entire population interviewed and then reinterviewed. Because only a sample of the population is interviewed and a fraction of that sample reinterviewed, these values can only be estimated in practice.

The symbol  $\pi_{00}$  represents the proportion in the *population* of those who would not report substance use in either the first or the second interviews;  $\pi_{01}$  represents the proportion of the population who would not report substance use in the first interview, but would report substance use in the second interview;  $\pi_{10}$  represents the proportion of persons in the population who would report substance use in the first interview, but not in the second interview; and  $\pi_{11}$  represents the proportion of the population who would report substance use in both the first and second interviews.

**Table 4.1. Basic 2x2 Contingency Table for Substance Use**

T1	T2 Nonuse (0)	T2 Use (1)	T2 Total
Nonuse (0)	$\pi_{00}$	$\pi_{01}$	$\pi_{0+}$
Use (1)	$\pi_{10}$	$\pi_{11}$	$\pi_{1+}$
Total	$\pi_{+0}$	$\pi_{+1}$	$\pi_{++} = 1$

Measures of agreement between T1 and T2 responses calculated for the key measures in this study include (weighted) percent reporting consistently, Cohen's kappa ( $\kappa$ ), and the index of inconsistency (*IOI*).

#### 4.1.1 Weighted Percent Reporting Consistently

An obvious measure of agreement is the *raw proportion reporting consistently*, defined as the proportion *within the population* of individuals who would have given the same response at T2 as they did at T1. In the case of a dichotomous variable, the raw consistent reporting rate is

$\pi_{00} + \pi_{11}$ , using the notation in Table 4.1. Weighted estimates of the raw percent reporting consistently are computed from the NSDUH sample because of the unequal selection probabilities and nonresponse adjustments. These estimates, given in percentages and referred to as the weighted percent reporting consistently, are provided as a basic measure of consistency. Weighted percent reporting consistently in its simplest form for a dichotomous variable such as substance use is defined as follows:

$$P_{Agree} = 100p_{Agree} = 100(p_{00} + p_{11}),$$

where  $p_{ij}$  is the weighted estimate of the population proportion,  $\pi_{ij}$ .

One weakness of this measure is that it does not account for chance agreement between responses. Chance agreement of a dichotomous variable is defined as

$$\pi_{Chance} = \pi_{0+}\pi_{+0} + \pi_{1+}\pi_{+1}.$$

See Agresti (2002) for a discussion of chance agreement. The next section introduces Cohen's kappa, which does account for chance agreement.

#### 4.1.2 Cohen's Kappa

Cohen (1960, 1968) introduced a measure of agreement (or "concordance") between measurements denoted by kappa ( $\kappa$ ) that does correct for chance agreement. This measure is the statistic most often used to assess interrater reliability of categorical variables. For this study, kappa was used to assess the agreement of responses given by the same respondent at two different times, T1 and T2. Because most of the measurements in the Reliability Study are dichotomous (binary), and because of its simpler form, this case is discussed first.

*Cohen's Kappa for Dichotomous Variables.* For dichotomous outcomes with levels 0 and 1, four possible combinations of responses are given by a respondent, where the first digit is the T1 response and the second is the T2 response: 00, 01, 10, and 11. The weighted estimate of the proportion of respondents falling into the combination  $i, j$  is denoted by  $p_{i,j}$ , corresponding to the population proportion  $\pi_{i,j}$  shown in Table 4.1. Similarly,  $p_{i+}$  and  $p_{+j}$  are the weighted estimates of the related marginal proportions. The raw proportion of agreement from Table 4.1 is given by  $\pi_{00} + \pi_{11}$ . As mentioned above, this measure does not account for chance agreement,

$$\pi_{Chance} = \pi_{0+}\pi_{+0} + \pi_{1+}\pi_{+1}.$$

Cohen's kappa, which corrects for chance agreement, is defined by

$$\kappa = \frac{\pi_{Agree} - \pi_{Chance}}{1 - \pi_{Chance}}.$$

Using the notation above,

$$\kappa = \frac{(\pi_{00} + \pi_{11}) - (\pi_{0+}\pi_{+0} + \pi_{1+}\pi_{+1})}{1 - (\pi_{0+}\pi_{+0} + \pi_{1+}\pi_{+1})}.$$

Replacing the population parameters in the definition of kappa above by their weighted estimates, the following is obtained:

$$\hat{\kappa} = \frac{(p_{00} + p_{11}) - (p_{0+}p_{+0} + p_{1+}p_{+1})}{1 - (p_{0+}p_{+0} + p_{1+}p_{+1})}.$$

The common approach to estimating the standard error for the kappa does not account for the sample design, thus potentially creating a downward bias. The sample for this study was drawn using a complex stratified multistage design. Feder (2006) developed a Taylor linearization (TL) variance estimation method that was implemented in a SAS<sup>®</sup>/SUDAAN<sup>®</sup> macro and used to derive all of the standard errors in this report (for details, see Section B.2 in Appendix B).

*Interpreting Cohen's Kappa.* Following Landis and Koch (1977, p. 165), this report used the following benchmarks for assessing the level of agreement based on the estimated kappa:

- poor agreement for kappas less than 0.00;
- slight agreement for kappas of 0.00 to 0.20;
- fair agreement for kappas of 0.21 to 0.40;
- moderate agreement for kappas of 0.41 to 0.60;
- substantial agreement for kappas of 0.61 to 0.80; and
- almost perfect agreement for kappas of 0.81 to 1.00.

The kappa measure is very sensitive to the prevalence rate. As a consequence, its behavior is sometimes regarded as paradoxical (for more information, see Section B.1 in Appendix B). Therefore, prevalence rates have been included in the tables containing kappa estimates in this report. These estimated prevalence rates were calculated from the Reliability Study's T1 responses. Furthermore, because of the dependence of kappa on the prevalence rates, direct comparison of kappa values between population groups having different prevalence rates may be inappropriate. However, when the prevalence rates are not very low or not very high, one can refer to the Landis and Koch (1977) benchmarks to assess the level of agreement.

*Cohen's Kappa Measure and Generalized Kappa (for Nondichotomous Variables).* Some of the categorical variables in the Reliability Study have more than two response levels. The generalized kappa is defined as

$$\kappa_w = \frac{\sum u_{ij}\pi_{ij} - \sum \sum u_{ij}\pi_{i+}\pi_{+j}}{1 - \sum \sum u_{ij}\pi_{i+}\pi_{+j}},$$

where  $u_{ij}$  is a "weight" given to the level of agreement between the potential responses  $i$  given at T1, and  $j$  given at T2. For example, one may define  $u_{ij}$  to be 1 if  $i = j$  (perfect agreement) and 0 otherwise. In this case,  $\kappa_w$  reduces to the usual Cohen's kappa. See also Agresti (2002).

In this report, a "relaxed agreement" criterion also was used where a difference no greater than  $k$  levels was allowed. This definition corresponds to the choice of  $u_{ij}$  to be 1 if  $|i - j| \leq k$ , 0 otherwise.

### 4.1.3 Index of Inconsistency

The index of inconsistency (*IOI*) is used in this report as a reliability measure for continuous and scale variables. It is defined by Pritzker and Hanson (1962) as

$$IOI = \frac{\sigma_R^2}{\sigma_T^2},$$

where  $\sigma_R^2$  is half the average squared difference between the potential T1 and T2 responses across the individuals in the population, and  $\sigma_T^2$  is the average total variance of potential responses across the population for both T1 and T2. (See Section B.3 in Appendix B for more discussion of the *IOI*.) Note that the *IOI* reflects lack of reliability and, as noted by Cochran (1977, p. 387), is analogous to  $1 - \phi$ , where  $\phi$  is the coefficient of reliability used in studying errors of measurement in psychology.

## 4.2. Models

The models discussed below were fitted to the data for several reasons:

1. To investigate the patterns of nonresponse, especially that of individuals who did complete the T1, but were nonrespondents at the T2 interview.
2. To investigate the relationship between the consistency of age at first use self-reporting and the time since first use (potentially affecting the recall bias).
3. To estimate the true prevalence of a NSDUH variable under a restricted model that nonetheless allows the weighted means from the interviews to be systematically biased.

### 4.2.1 Response Propensity Models

Logistic regression models were fitted where the response variable was 1 if an interview (T1 or T2) was completed and 0 otherwise. The independent variables were "age," "gender," "race/ethnicity," "region," and census tract data: "population density," "percent families below poverty level," "percent housing units rented," "percent population with bachelor's, graduate, or professional degree," "median household income," and "urbanicity." For the T2 response model, "reported mental health problems at T1" and "reported cocaine or heroin use at T1" also were

included as covariates to see whether these stigma-carrying responses affected the T2 response propensity.

#### **4.2.2 Logit Model to Assess Dependence of Reliability of Age at First Use and Duration of Recall**

It is of interest to know whether the time since first use affects the consistency of self-reporting of the age at first use. If longer time users are less precise in this respect, a recall bias may be present. Thus, logistic regression models were fitted where the response variable was defined as 1 when the age at first use reported at T1 matched that reported at T2. Another version of the logistic regression models was fitted where the response variable was defined as 1 when the age at first use reported at T1 and the age at first use at T2 were the same, or differed by 1 year. The independent variables were the "years since first use" (calculated using the T1 data), "gender," "education," "race/ethnicity," and "whether both interviews were administered by the same field interviewer." Similar models also were fitted in which age (at T1) also was included as an independent variable.

#### **4.2.3 Hui-Walter Modeling**

So far, the possibility has not been ruled out that although the survey responses for an individual may vary from T1 to T2, the weighted estimates based on the two sets of data are free of systematic bias. What if that were not the case? In particular, what if the false negative rate were larger than the false positive rate? For a question such as past year marijuana use, this means that a larger number of population users in the previous year would answer incorrectly to T1 (and to T2) than the number of nonusers would.

To investigate this question, a modeling approach introduced by Hui and Walter (1980) was adopted and applied to the reinterview dataset for past year marijuana use. The methodology required the population to be divided into two groups (males and females were chosen) such that the true prevalences were different within each of the groups but the false positive and false negative rates were not. Although the analysis allowed some testing of these restrictive assumptions, they remained fairly strong. The methodology used is discussed in more detail in Section 6.5 in Chapter 6.





## 5. Descriptive Survey Statistics

This chapter presents the basic descriptive statistics of the sample respondents, associated response rates, interview timing data, and time between the initial interviews (T1) and the follow-up interviews or reinterviews (T2) for the 2006 National Survey on Drug Use and Health (NSDUH) Reliability Study. Its purpose is to provide the reader with an overview of the survey respondents and the interviewing parameters for the Reliability Study.

### 5.1. Survey Sample and Respondent Characteristics

The sample for the Reliability Study included persons aged 12 years or older living in households in the 48 contiguous States who chose to complete the T1 interview in English. This sample excluded all persons living in noninstitutional facilities, such as shelters, rooming houses, fraternity and sorority houses, and dormitories. Persons living in households in which two respondents were selected for the main study also were excluded. A total of 4,544 persons were selected for the Reliability Study. Of that total, 3,136 persons (69 percent) completed both the T1 and T2 interviews. Table 5.1 shows the distribution of the T1 and T2 samples by age group, gender, race/ethnicity, and the same versus different field interviewer (FI) substudy.

**Table 5.1. Sample and Response Distribution: 2006 NSDUH Reliability Study**

Domain	T1		T1		T2		T2	
	Number Selected	Percent Selected of Total	Number Complete	Percent Complete of Total	Number Eligible	Percent Eligible of Total	Number Complete	Percent Complete of Total
<b>Total</b>	4,544	100.0	3,634	100.0	3,516	100.0	3,136	100.0
<b>Age</b>								
12-17	1,240	27.3	1,069	29.4	1,044	29.7	976	31.1
18-25	1,335	29.4	1,094	30.1	1,049	29.8	948	30.2
26 or Older	1,969	43.3	1,471	40.5	1,423	40.5	1,212	38.6
<b>Gender</b>								
Male	2,122	46.7	1,674	46.1	1,613	45.9	1,419	45.2
Female	2,422	53.3	1,960	53.9	1,903	54.1	1,717	54.8
<b>Race/Ethnicity</b>								
Hispanic	518	11.4	431	11.9	324	9.2	280	8.9
Non-Hispanic								
White	3,163	69.6	2,498	68.7	2,492	70.9	2,229	71.1
Non-Hispanic								
Black	597	13.1	486	13.4	483	13.7	434	13.8
Other	266	5.9	219	6.0	217	6.2	193	6.2
<b>Same versus Different Field Interviewer (FI)</b>								
Same FI	2,964	65.2	2,339	64.4	2,271	64.6	2,070	66.0
Different FI	1,580	34.8	1,295	35.6	1,245	35.4	1,066	34.0

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

## 5.2. Response Rates

Response rates were calculated for the first interview (T1) and for the reinterview (T2), both weighted and unweighted, for the entire study sample, and for the same versus different FI subsamples (for details, see Section 3.1.2 in Chapter 3) and by age group, gender, and race/ethnicity. Table 5.2 presents these response rates. The response rates at T2 are conditional on completion of the T1 interview.

The analytic weights for this study were calibrated to the national control totals. Thus, weighted estimates in this report are representative of the population. The control totals used in the poststratification in the Reliability Study were estimated from the census population estimates, excluding the population from Alaska and Hawaii, the population of institutional group quarters, and the population that does not speak English. To obtain the estimated control totals, the proportion of NSDUH respondents interviewed in nongroup quarters and in English was calculated using the pooled 2004, 2005, and 2006 NSDUH data for each demographic domain formed by intersecting age group, race, Hispanicity, and gender. These domain-specific proportions then were multiplied by the 2006 population estimates provided by the Census Bureau.

The weighted response rates for the T1 interview (74.9 percent for the total) are slightly better than the weighted response rates for the full 2006 NSDUH (74.2 percent for the total) (Office of Applied Studies [OAS], 2007). The unweighted response rates for T2 were slightly lower than the design parameter response rates based on the pretest.

## 5.3. Interview Timing Data

The two main measures of the duration of the interview are the total interview time and the duration of the audio computer-assisted self-interviewing (ACASI) part of the survey. These duration measures were available for both interviews, T1 and T2. Cases whose timing data for either the T1 or T2 interview were questionable were removed. Such cases included those where the time to complete either the full interview or the ACASI part of the interview was negative because of an incorrect time on the laptop on which the interview was conducted, cases where the interview was incomplete, and cases where one of the two interviews had at least one break-off. Cases also were removed if the time to complete the interview or the ACASI part was outside the 2 or 98 percentile at T1 or T2. Table 5.3 shows summary statistics on these measures for the full study sample and for the same versus different FI subsamples.

Table 5.4 shows summary statistics on the difference in the total interview and the ACASI durations, including *p* values for testing the hypothesis of the mean difference being zero.

The duration of the interview at T1 and at T2 are obviously related. Table 5.5 shows summary statistics on the correlations between the durations (also see Figure 5.1).

**Table 5.2. Final Unweighted and Weighted T1 and T2 Response Rates: 2006 NSDUH Reliability Study**

<b>Domain</b>	<b>T1 Number Selected</b>	<b>T1 Number Selected Complete</b>	<b>T1 Unweighted Response Rate</b>	<b>T1 Weighted Response Rate</b>	<b>T2 Number Eligible</b>	<b>T2 Number Eligible Complete</b>	<b>T2 Unweighted Response Rate</b>	<b>T2 Weighted Response Rate</b>
<b>Total</b>	4,544	3,634	80.0	74.9	3,516	3,136	89.2	85.6
<b>Age</b>								
12-17	1,240	1,069	86.2	87.4	1,044	976	93.5	94.0
18-25	1,335	1,094	82.0	81.6	1,049	948	90.4	89.4
26 or Older	1,969	1,471	74.7	73.0	1,423	1,212	85.2	84.4
<b>Gender</b>								
Male	2,122	1,674	78.9	75.2	1,613	1,419	88.0	84.2
Female	2,422	1,960	80.9	74.6	1,903	1,717	90.2	86.9
<b>Race/Ethnicity</b>								
Hispanic	518	431	83.2	82.6	324	280	86.4	80.1
Non-Hispanic White	3,163	2,498	79.0	74.4	2,492	2,229	89.5	86.0
Non-Hispanic Black	597	486	81.4	76.3	483	434	89.9	85.5
Other	266	219	82.3	63.6	217	193	88.9	88.0
<b>Same versus Different Field Interviewer (FI)</b>								
Same FI	2,964	2,339	78.9	74.7	2,271	2,070	91.2	89.3
Different FI	1,580	1,295	82.0	75.2	1,245	1,066	85.6	78.3

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 5.3. Summary Statistics on Interview Length, in Minutes: 2006 NSDUH Reliability Study**

<b>Measure</b>	<b><math>n^1</math></b>	<b>Mean</b>	<b>Minimum</b>	<b>Median</b>	<b>Maximum</b>
<b>Total Interview</b>					
T1	2,775	58.1	35.1	56.4	100.8
T2	2,775	54.5	31.0	52.2	95.5
T2 Same Field Interviewer (FI)	1,831	54.5	31.0	53.2	94.1
T2 Different FI	944	56.2	31.0	52.0	95.5
<b>Audio Computer-Assisted Self- Interviewing (ACASI) Portion</b>					
T1	2,775	38.3	17.9	35.9	76.3
T2	2,775	31.2	12.6	29.2	65.8
T2 Same FI	1,831	31.4	12.6	30.5	65.8
T2 Different FI	944	32.0	12.6	28.9	65.7

<sup>1</sup> Cases whose timing data were questionable were removed.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 5.4. Difference in Time to Complete Interview (in Minutes) at T1 and T2, with T2 by Same and Different FI, Including *p* Value for Test of Mean of T1 – T2 Being Zero: 2006 NSDUH Reliability Study**

Measure	<i>n</i> <sup>1</sup>	Mean	Minimum	Median	Maximum	<i>p</i> Value
<b>Total Interview</b>						
T1 – T2	2,775	3.7	-49.9	3.7	49.5	0.0000
T1 – T2 Same Field						
Interviewer (FI)	1,831	3.7	-44.8	3.3	41.0	0.0000
T1 – T2 Different FI	944	3.1	-49.9	4.0	49.5	0.0009
<b>Audio Computer-Assisted Self-Interviewing (ACASI) Portion</b>						
T1 – T2	2,775	7.1	-38.1	6.6	45.2	0.0000
T1 – T2 Same FI	1,831	6.9	-30.5	6.3	45.2	0.0000
T1 – T2 Different FI	944	7.2	-38.1	7.6	43.0	0.0000

<sup>1</sup> Cases whose timing data were questionable were removed.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 5.5. Correlations between T1 and T2 Timing Data: 2006 NSDUH Reliability Study**

Measure	Time to Complete Interview	Time to Complete Audio Computer-Assisted Self-Interviewing (ACASI) Portion
Total Sample	0.6131	0.7229
Same Field Interviewer (FI)	0.6201	0.7329
Different FI	0.5845	0.6752

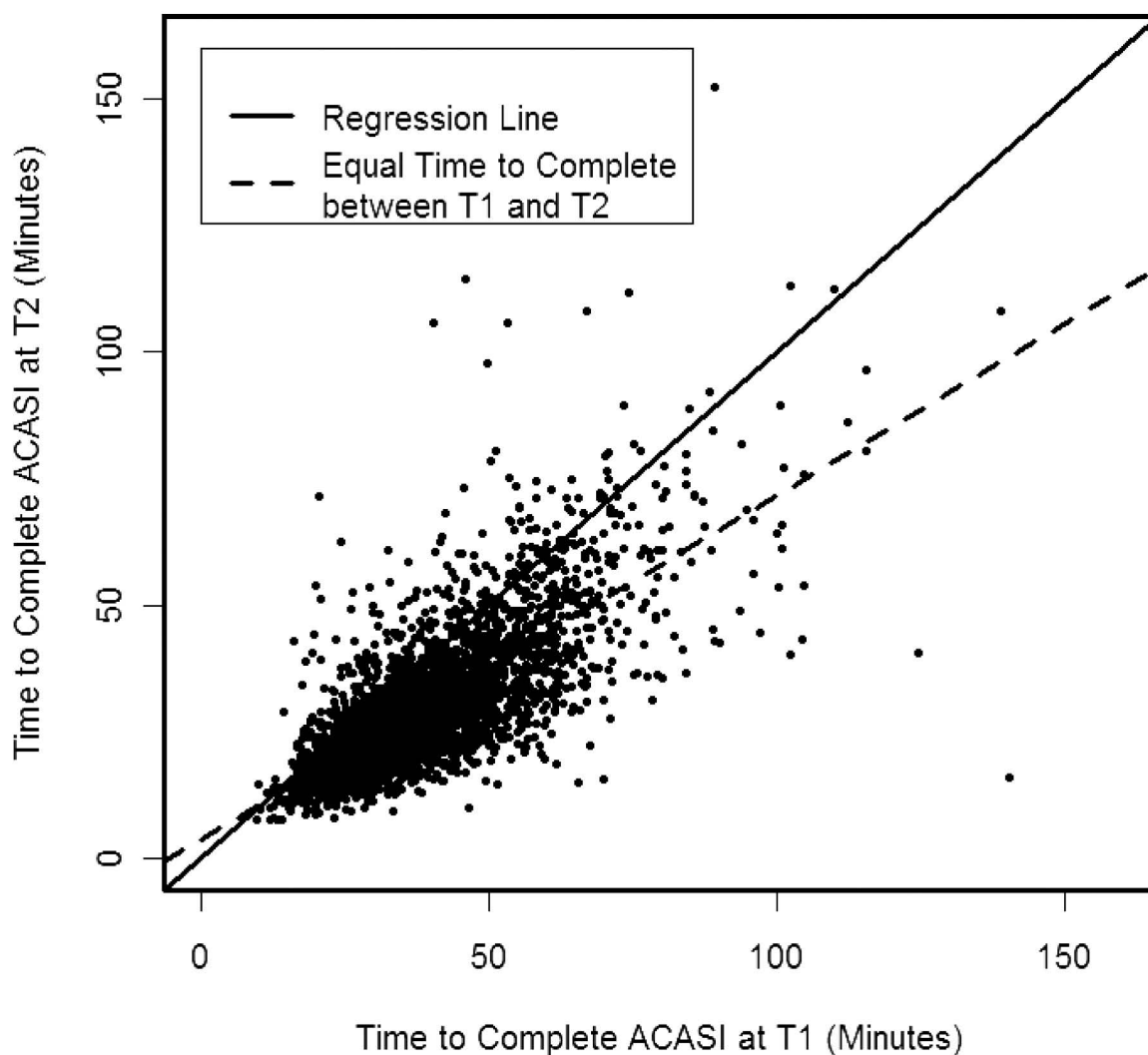
Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

## 5.4. Time between T1 and T2 Interviews

As a rule, the T2 interview was to be conducted between 5 and 15 days after the T1 interview. There were a few exceptions to this rule, including five cases where the difference (referred to as the "lag") was less than 5 days and two cases where it was more than 15 days. The mean lag was 8.4 days, and the median was 7 days.

Table 5.6 provides the number of cases by the number of days lag between the T1 and T2 ACASI interviews. An analysis of the effects of lag time on response agreement for several measures appears in Chapter 8.

**Figure 5.1. Relationship between the T1 and T2 Durations of the Audio Computer-Assisted Self-Interviewing (ACASI) Interviews: 2006 NSDUH Reliability Study**



Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 5.6. Distribution of Completed Cases by the Lag Time between T1 and T2 Audio Computer-Assisted Self-Interviewing (ACASI) Interviews in Days: 2006 NSDUH Reliability Study**

Lag (Days)	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Number of Completed Cases	3	2	516	389	721	274	270	193	157	159	161	219	70	1	1

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).



## 6. Reliability Findings

The tables presented at the end of this chapter constitute the main body of statistical data from the 2006 National Survey on Drug Use and Health (NSDUH) Reliability Study. They provide reliability measures on many of the variables derived from the questionnaire responses after light editing of some variables for consistency and ease of analysis (see Section 3.8). The estimated reliability measures shown in the tables are as follows:

- weighted percent reporting consistently (see Section 4.1.1);
- Cohen's kappa (see Section 4.1.2) for dichotomous variables;
- generalized kappa using a "relaxed agreement" criterion (see Section 4.1.2) for other categorical variables, including scale variables; and
- index of inconsistency (*IOI*, see Section 4.1.3) for continuous and scale variables (thus, for scale variables, the *IOI* is shown in addition to the generalized kappa).

Note that the last measure reflects the lack of consistency (i.e., its complement,  $1 - IOI$ , increases with the response's consistency). It also should be noted that each table presented in this chapter does not contain every one of the reliability measures. Tables that contain only dichotomous variables do not show the generalized kappa or the *IOI*. Tables that have a mix of dichotomous, categorical, and continuous variables may include all of the reliability measures, but an "N/A" (not applicable) will be shown for those measures that do not correspond to the variable type.

Because of the sensitivity of Cohen's kappa to prevalence rates, weighted prevalence rates were estimated using the initial interview, or T1, responses and were included in the tables for each of the dichotomous variables. In principle, weighted prevalence rates for each level of the other categorical variables could have been presented for the generalized kappa estimates. However, this would have been difficult to present in an easily understandable way in the tables and perhaps would not have been as informative as in the dichotomous case. Suppression of kappa estimates for the case in which the prevalence rate is very low or very high is described in Section B.4 in Appendix B.

Also included in the tables are the number of respondents with nonmissing data at both interviews and the weighted percent missing. These tables provide a brief summary of the response consistency of data pertaining to each variable.

Generally, the tables in this chapter are presented in the order in which the variables they contain appear in the questionnaire. However, there is some deviation from this general approach. In some cases, variables were grouped into tables on the basis of their type (dichotomous, ordinal, other categorical type, or continuous) or topic in spite of their appearance at different places in the questionnaire.

Some of the questions in the NSDUH instrument have ordinal responses with many levels, and a small variation in the responses may be tolerated. One example of this type of variable is the number of persons living in a household, and another is the four-level composite annual family income variable. Therefore, a relaxed agreement criterion also was used and reported in some tables, where response differences no greater than a given threshold were allowed. The threshold allowed under the relaxed agreement criterion is specified in the footnotes for the respective variables in each applicable table.

## 6.1. Overview of Analyses

The examination of response reliability for the 2006 NSDUH questionnaire items and resulting variables is primarily based on Cohen's kappa (see Section 4.1.2). Recall that Landis and Koch (1977) characterized kappa values over 0.8 as "almost perfect" and those in the 0.6 to 0.8 range as "substantial."

For continuous and scale-type variables, the *IOI* is presented (see Section 4.1.3). As noted previously, the *IOI* often is interpreted as "low" (meaning high consistency) if it is less than 0.20, "moderate" if between 0.20 and 0.50, and "high" if over 0.50. See Section B.3 in Appendix B for the relationship between the *IOI* and the kappa and characterization of *IOI* values.

Most kappa values estimated in the tables presented in this chapter are over 0.60, which indicates substantial reliability. However, in some instances, the estimated kappa values are lower, indicating that the variables to which they relate are less reliable (for the remainder of this section, the modifier "estimated" is dropped from estimated values for lucidity). There are two exceptions to this conclusion:

1. Differences between the responses in the two interviews may be due to a true change occurring in the few days between the interviews.
2. When the prevalence rate is very close to 0 percent or to 100 percent, a very small number of inconsistent respondents may result in a low kappa.

It also can be seen that questions that ask for an attitude or opinion or an intent to some kind of action show less reliability. With the enormous amount of and access to communication media, an individual's opinions and attitudes are constantly being influenced; thus, changes in responses to these types of questions over a short period of time could be expected.

Reliability results for selected 2006 NSDUH questionnaire variables are presented in Section 6.2. Also presented in this chapter are the results of the modeling of the questionnaire responses: response propensity model analyses (Section 6.3), analyses of age at first use reliability and duration of recall (Section 6.4), and Hui-Walter modeling (Section 6.5).

## 6.2. Reliability Statistics, by Type of Variable

Findings are presented in Tables 6.1 to 6.47 for the following types of variables: substance use, perceived risk and availability of substances, substance dependence and abuse, illegal behaviors and marijuana acquisition, timing of substance use, substance use treatment,



health conditions and health care, mental health treatment, social behavior and attitudes, mental health problems, alcohol use, and demographic characteristics.

### 6.2.1 Substance Use

Tables 6.1 and 6.2 present reliability statistics for lifetime and past year use of the substances asked about in the core section of the NSDUH questionnaire. Kappa values in Table 6.1 for the lifetime use variables were all 0.70 or better. Most of them were above 0.80, which is characterized as "almost perfect." The kappa values for the past year variables in Table 6.2 showed very good reliability, but were not as impressive as the lifetime use values. Of the substances with nonsuppressed kappa values, nearly half of them were greater than 0.80. All of the other kappas were greater than 0.60, which is characterized as "substantial" agreement. Most of the substances with the kappa values of less than 0.80 also had low prevalence, which affects the kappa (see Appendix B).

Table 6.3 presents reliability statistics for reported age at first use of individual substances both for exact age agreement at both interviews and for approximate agreement of age where a difference of 1 year of age is still considered an agreement. As might be expected because of having to recall more detailed information, these variables showed less reliability than the questions about lifetime and past year use. For exact agreement, the kappas ranged from 0.21 to 0.58. Not surprisingly, for the approximate agreement, the kappas ranged from 0.47 to 0.85. In fact, nearly 75 percent of the kappas for the approximate agreement were greater than 0.60, which indicates "substantial" agreement. The *IOI* values generally gave similar indications of reliability as the kappas. However, in those cases where differences between the ages given at the two interviews were greater than 1 year, the kappa values were lowered (both under exact and approximate agreement criteria). However, the *IOI* was still good because the differences were not great relative to the range of the corresponding age at first use.

Table 6.4 presents reliability statistics for questions about which substance (alcohol, marijuana, or cigarettes) was used first when the age at first use for at least two of the substances was reported to be the same. The kappa values were inconsistent in this table, with some very high and some low. The small number of respondents for these items was most likely contributing to this inconsistency.

Table 6.5 contains variables that attempt to quantify substance use over a period of time. The reliability statistics were computed both for exact agreement and for approximate agreement. The variable "number of months from first alcohol use to the time of the initial interview" was asked only of persons whose age at first use was either their current age or 1 year younger than their current age. This was a fairly short recall period for what is likely a very significant event in a person's lifetime. Thus, the kappas for exact agreement and for approximate agreement were both in the "substantial" agreement range. Unfortunately, nearly half of the persons eligible for this question did not provide sufficient information to allow the number of months to be calculated, which raises doubts about the validity of the kappa. The *IOI* for this variable was also relatively high, indicating that the reliability was questionable. Kappas for the "number of days used in the past year" for both exact agreement and approximate agreement were under 0.50 and in the ranges characterized as "fair" or "moderate." The corresponding *IOI*

values for this variable were not sufficiently small as to bring into question the findings from the kappas.

Table 6.6 contains variables related to the use of blunts, defined as cigars with marijuana in them. Reliability statistics were computed both for exact and approximate agreement. As was the case for lifetime and past year use of other substances, lifetime and past year use of blunts had kappas greater than 0.60 for those cases that were not suppressed. Also, in line with the pattern already seen with other substances, age at first use had a low kappa under the exact agreement requirement, but improved to over 0.60 under approximate agreement of no more than 1 year difference between what was reported at the initial and follow-up interviews.

Table 6.7 presents reliability statistics for lifetime and past year use of specific substances. All of the kappas that were not suppressed except for one variable ("lifetime use of nonprescription cough or cold medicine") were greater than 0.60 ("substantial" agreement). A note of caution when reviewing this table is that all of the weighted prevalence rates were less than 3 percent.

### **6.2.2 Perceived Risk and Availability**

Table 6.8 presents reliability statistics for risk and availability questions for the full set of response options and for the dichotomized response options as specified in the footnotes of the table. Most kappas were in the lower ranges, below 0.60. As expected, the kappas improved when the response options were dichotomized. Most of the questions asked for the respondent's opinion or judgment, which was influenced daily through personal interactions and communications media. Thus, these types of questions were subject to being less reliable. The questions that were more factual, such as wearing a seatbelt when riding in or driving a car, were much less prone to regular fluctuations. The kappas in Table 6.8 for these questions ranged from 0.72 to 0.78, which are characterized as "substantial" agreement. The kappas for the seatbelt questions improved to over 0.80 ("almost perfect" agreement) when the response options were dichotomized.

### **6.2.3 Substance Dependence and Abuse**

Tables 6.9 to 6.12 present reliability statistics for dependence and abuse indicators and their component questions for alcohol, marijuana, cocaine, and pain relievers, respectively. The kappas in Table 6.9 for the alcohol dependence, abuse, and dependence or abuse indicators were all over 0.60. The kappas for the individual criteria making up the dependence and abuse indicators were not as consistently high. In Table 6.10, the kappas for the marijuana dependence and abuse indicators were not as high as the kappas for alcohol measures. Interestingly, the kappas for the individual criteria for the marijuana dependence and abuse indicators were generally higher than those for the respective overall indicator. These indicators were also generally higher than those for alcohol dependence and abuse. The kappas in Table 6.11 for cocaine dependence and dependence or abuse were similar to the corresponding kappas for marijuana dependence and dependence or abuse. However, the kappa for cocaine abuse was very high, 0.84. The kappas for the individual cocaine dependence criteria were generally low. On the other hand, the kappas for the individual cocaine abuse criteria were greater than 0.80, except for

one item that was 0.66. In Table 6.12, only the kappa for the pain relievers dependence indicator was greater than 0.60.

Table 6.13 presents reliability statistics for dependence, abuse, and dependence or abuse variables for illicit drugs and/or alcohol. The kappas were generally greater than 0.60, except for dependence on illicit drugs and alcohol and dependence on or abuse of illicit drugs and alcohol.

#### **6.2.4 Illegal Behaviors and Marijuana Acquisition**

Table 6.14 presents reliability statistics for variables regarding illegal behaviors and variables about legal issues regarding marijuana. The kappas in this table were nearly all greater than 0.60. The one variable with a kappa of 0.07 was for the question about the legal penalty for marijuana possession. This is not surprising because the weighted percent of missing data for this item was 38 percent, indicating that a substantial proportion of respondents did not know the penalty; indeed, most of the missing data were related to "don't know" responses to this item.

Table 6.15 presents reliability statistics for how marijuana users last obtained the marijuana they used. The kappas for these variables were all very high. For the two scale variables in this table, the *IOI* value did not refute the reliability shown by the kappa.

#### **6.2.5 Timing of Substance Use**

In Table 6.16, the kappas were all 0.75 or greater for use of substances in the year prior to the interview, indicating "substantial" (or better) agreement.

Table 6.17 presents reliability statistics for the age at last use of a substance. This table shows statistics for both exact age agreement and for ages at the two interviews being different by no more than 1 year. The kappas for exact agreement ranged from 0.31 to 0.62, with most of the kappas less than 0.50. The kappas improved for approximate agreement, with a range from 0.37 to 0.80 and with most of the kappas greater than 0.60. The one kappa of 0.37 for age at last use of tranquilizers was confirmed by the *IOI* value of 0.27.

Table 6.18 presents reliability statistics for the number of months since last use of cigarettes, cigars, alcohol, and marijuana. These questions were asked only of persons whose reported last use of the substance either at their current age or 1 year younger than their current age. This table presents kappas for both exact and approximate agreement, which is defined as the number of months since last use being different by no more than 1 month for the two interviews. The exact agreement kappas ranged from 0.41 to 0.64, and the approximate agreement kappas ranged from 0.55 to 0.74. The *IOI* values showed only moderate consistency for these variables. One item of concern for this table is its somewhat large weighted percent missing values for each variable.

#### **6.2.6 Substance Use Treatment**

Table 6.19 presents reliability statistics for variables associated with receipt of treatment for alcohol or drug use. All kappas were over 0.80 with one exception, "enrolled in treatment program on October 1st of last year," which had a kappa of 0.42. A possible explanation of this is that there was a misunderstanding of what "enrolled in" means. The term "enrolled in" may

have been sometimes ambiguous to a subject in the sense that he or she went to treatment regularly but did not consider that attendance as being "enrolled" in a program. It may also be that someone who was in an ongoing program that started prior to October 1st interpreted this question as enrolling on October 1.

### **6.2.7 Health Conditions and Health Care**

Tables 6.20 and 6.21 present reliability statistics for past year health conditions and for hospital treatment of some conditions. The kappas in Table 6.20 were mostly greater than 0.65, except for three conditions ("pneumonia," "sexually transmitted disease," "sinusitis") that had kappas around 0.55. These variables' fairly low prevalence rates may have affected the value of the kappas.

Table 6.21 presents reliability statistics on the frequency of past year health care both for exact and approximate agreement. Two variables in Table 6.21 ("number of times been treated in an emergency room" and "number of nights stayed in a hospital overnight or longer") were follow-up questions to the two corresponding questions in Table 6.20. The kappas for these variables' exact agreement were 0.57 and 0.65, which rose to over 0.70 for the number of nights stayed in a hospital when the agreement criteria were relaxed. The *IOI* values here showed fairly high inconsistency, which was counter to what the kappa values showed.

Table 6.22 presents reliability statistics for lifetime health conditions. The kappas were all 0.68 or greater, except for "stroke" (0.59) and "ulcer or ulcers" (0.53). The prevalence for both these conditions was less than 3 percent, which may have contributed to lower than expected kappa values.

### **6.2.8 Mental Health Treatment**

Tables 6.23 to 6.25 present reliability statistics for various mental health treatment variables. The very small number of respondents for the inpatient treatment variables made the kappa values unreliable, so those results were not presented. Except for six variables, the kappas for the outpatient treatment variables in Table 6.23 were greater than 0.65. Four of the six variables with low kappas were location of treatment variables: "outpatient mental health clinic or center" (0.43), "treatment at a doctor's office – not part of a clinic" (0.52), "partial day hospital or day treatment program" (0.03), and "outpatient medical clinic" (0.02). The other two low kappa variables were source of payment variables: "Medicare" (0.50) and "other private source" (0.00).

Table 6.21 (partially discussed earlier) also presents reliability statistics for the number of visits made to a private therapist for persons aged 12 or older who reported treatment at a private therapist. Statistics are shown both for exact and approximate agreement when the number of visits reported in the two interviews was within five visits. The weighted percent reporting consistently for exact and approximate agreement was 42 and 88 percent, respectively. The corresponding kappas were 0.38 and 0.86. Thus, these statistics did not show good reliability for this variable. The *IOI* statistic confirmed the approximate agreement kappa.

Table 6.24 presents reliability statistics for unmet need for mental health treatment and the reasons for not receiving such treatment. The kappas were generally very good (over 0.65),

except for three variables: "health insurance does not cover any mental health treatment/counseling" (0.16), "did not know where to go for services" (0.17), and "some other reason" (0.47). The weighted percent reporting consistently for the variable "some other reason" was only 73.5 percent, which confirmed that this question was not highly reliable.

Table 6.25 presents reliability statistics for alternative mental health treatment and types. The kappas for all of these variables were 0.60 or greater.

### **6.2.9 Social Behavior and Attitudes**

Table 6.26 presents reliability statistics for social behavior and attitude variables among persons aged 18 or older. Statistics were computed both for exact and approximate agreement that was defined for each variable. The kappas for exact agreement ranged from 0.43 to 0.77, with about half of the kappas less than 0.60. The kappas for approximate agreement ranged from 0.56 to 0.93, with only one kappa less than 0.60. The *IOI* values for these variables were generally in line with the kappas, but did not show quite as strong reliability as the kappas.

Tables 6.27 to 6.29 present reliability statistics for social attitudes and behaviors of respondents aged 12 to 17 years old. In general, the kappas were less than 0.60, although a few kappas were 0.75 or greater. Because these types of variables could be influenced daily by the communications media and personal interactions, it was not surprising to find that they were less reliable than the more factually based variables.

### **6.2.10 Mental Health Problems**

Table 6.30 presents reliability statistics for serious psychological distress (SPD). The kappa for the dichotomous SPD indicator, derived additively with a predetermined cutoff from the individual K6 item responses, was 0.64. The kappa for the K6 score, which ranged from 0 to 24, was weak (0.21) when exact agreement was required. However, when the K6 scores were allowed to be no more than 3 points different between the two interviews, the kappa increased to 0.63. The kappas for each of the six items used to determine the K6 score were all 0.60 or greater, except for the kappa for how often a person felt nervous, which was 0.54. When kappa was computed under approximate agreement conditions for these six items, the kappa values were all 0.70 or greater. The *IOI* values for these variables led to similar conclusions as the kappas.

Tables 6.31 and 6.32 present reliability statistics for the depression variable among those respondents aged 18 or older. In Table 6.31, the variable indicating that a person had a major depressive episode (MDE) at some time in his or her life had a kappa of 0.67. The nine variables used to determine lifetime MDE did not have consistently high kappas. Three of the variables had kappas of 0.65 or greater. The variable indicating MDE in the past year had a kappa of 0.52. The rest of the variables in Table 6.31 generally had kappas greater than 0.60. However, there were five variables in the types of professionals seen or talked to for depression that had kappas ranging from 0.10 to 0.46.

Table 6.32 shows adult depression reliability statistics both for exact and approximate agreement. Kappas under exact agreement for these variables were all less than 0.60, whereas kappas under approximate agreement were all 0.65 or greater except for the variable "number of

days totally unable to work/do normal activities in the past year" that had a kappa of 0.47. This also had a low weighted percent reporting consistently of 58.9 percent. The *IOI* value for this variable was 0.05, indicating high consistency. The *IOI* values for the other variables were not as strong as the corresponding kappa values, but were not such that they refuted the general conclusions from the kappa values.

Table 6.33 presents reliability statistics for mental health variables for respondents aged 12 to 17. Only three variables had kappas in this table that were 0.60 or greater: (a) past year receipt of school counseling for emotional or behavioral problems from a "school counselor, school psychologist, or regular meetings with teachers"; (b) whether the respondent had "ever been in jail or detention center"; and (c) and whether the respondent had "ever been in foster care."

Tables 6.34 and 6.35 present reliability statistics for depression variables among respondents aged 12 to 17. Nine of the sixteen variables in Table 6.34 had kappa values less than 0.60. Most of these variables also had a relatively low weighted percent reporting consistently. Table 6.35 presents reliability statistics for depression variables among youths who were classified as having MDE at some time in their life or who met the suicide ideation criterion. Statistics are shown for both exact and approximate agreement. Only the kappa value for approximate agreement of age at first MDE episode was over 0.60. The weighted percent reporting consistently and the *IOI* for these variables were also in accord with the kappa values.

### **6.2.11 Alcohol Use**

Table 6.36 presents reliability statistics for alcohol consumption variables both for exact and approximate agreement among persons aged 12 or older. For exact agreement, two variables stood out with kappa values greater than 0.70: "ever had 5+ drinks on the same occasion" and "ever had 4+ drinks on the same occasion."

### **6.2.12 Demographic Characteristics**

Table 6.37 presents reliability statistics for the variables on the number of people living in the household and family income. The exact agreement kappa values for these variables were greater than 0.75, and the kappa values for approximate agreement were greater than 0.95. The *IOIs* for these variables confirmed their strong reliability.

Tables 6.38 to 6.40 present reliability statistics for the health insurance coverage variables. In Table 6.38, all but one of the kappas was greater than 0.65. The exception was "covers treatment for mental or emotional problems," which had a low kappa of 0.45. Table 6.39's two variables ("number of months without insurance/coverage in the past year" and "time since last had health care coverage") had kappas of 0.42 and 0.57 under exact agreement, rising to 0.58 and 0.86, respectively, under approximate agreement. Table 6.40 presents reliability statistics for the variables on the main reason that respondents aged 12 or older said that they were no longer being covered by health insurance; these respondents were currently without health care coverage, but had had it in the past. About half of the items in Table 6.40 had low kappa values, and half had kappa values greater than 0.60. A likely contributor to some of the

low kappa values in this table were the associated weighted prevalence rates that were below 3 percent.

Table 6.41 presents reliability statistics for types of income variables. The kappas were generally high for the variables in this table. Two kappas were just below 0.60 and related to prevalence rates of 3 percent or less. However, one item, on the interest, dividend income, or other income from family members, had a kappa of 0.45. It is not surprising that respondents did not have enough information about other family members' income from these sources to reliably report it.

Tables 6.42 and 6.43 present reliability statistics for the demographic variables. One would have expected these variables to show high reliability, and that is generally what was seen. However, in Table 6.42, there was one exception, "lived in the United States at least 1 year," that had a kappa of 0.31. An investigation into this unexpected outcome showed that the kappa was due to a single respondent with a very large weight; the unweighted kappa for this same variable was 0.82. In Table 6.43, there were three exceptions to high reliability. Two of them, "number of days of school missed because sick/injured" and "number of days school skipped because didn't want to go," each had a kappa value of 0.47 for exact agreement. Their corresponding kappa values for approximate agreement were 0.68 and 0.71, respectively. The third item with low reliability in Table 6.43, "what written policy covers," was a question asked of persons who were employed and who reported that their employer had a written policy regarding alcohol or drug use to determine if the policy covered only alcohol or drugs or covered both alcohol and drugs. The kappa value for exact agreement on this item was only 0.31.

Table 6.44 presents reliability statistics for the ethnicity, race, and marital status variables. The kappa values for all but one of the variables in this table were 0.95 or greater; the exception was for "American Indian or Alaska Native" (0.54). This low kappa value was likely affected by a low weighted prevalence rate of 1.1 percent.

Table 6.45 presents reliability statistics for several noncore demographic categorical variables, such as "State of residence," "education," "employment," and "country or U.S. Territory of birth." The kappa values for the variables in this table were all 0.65 or greater.

Table 6.46 presents reliability statistics for industry and occupation codes at a highly specific 4-digit code level and a more general 2-digit code level. All of the kappa values for the variables in this table were 0.70 or greater.

Table 6.47 presents reliability statistics for variables on the nonmedical use of specific pain relievers in the lifetime of persons aged 12 or older. Because of the low weighted prevalence rates of many of these variables, the kappa values were suppressed. For those variables with kappa values shown, the values were all greater than 0.65.

### **6.3. Response Propensity Model Findings**

Nonresponse is an important issue in almost any survey. One of the reasons for nonresponse is the burden on the respondent. Another hypothesized reason is the degree of personal sensitivity of the subject of the survey, in this case, illegal drug use history and mental

health problems. To minimize nonresponse, various procedures were used, including offering the \$50 incentive and being flexible about the time for the reinterview. These procedures resulted in a high (85.6 percent) weighted T2 response rate. However, it is important nonetheless to account for nonresponse due to the bias that it may induce.

Although nonresponse bias is always a concern, some of its effect is ameliorated by weighting. Recall that the Reliability Study sample was embedded within the main study sample in the 2006 NSDUH. Therefore, the weight for the Reliability Study sample was a product of the fully adjusted NSDUH main study analysis weight (ANALWT) and the weight used to select NSDUH respondents for reinterviewing in the Reliability Study, with adjustments for nonresponse and poststratification. The nonresponse bias is thereby accounted for by means of these adjustments. However, nonresponse adjustments are not perfect. It may still be of interest to know which factors may have contributed to nonresponse at either the initial interview (T1) or the reinterview (T2). For example, respondents who had reported illegal substance use at T1 may have been less likely to respond at T2. (See the response rates for the T1 and T2 interviews presented in Section 6.2.)

Tables 6.48 to 6.50 present results from the logistic regression models in which the dependent variable was defined as 1 if the selected individual responded (i.e., to the initial interview in some of the models and to both interviews in the rest of the models), and 0 otherwise. The models included the following independent variables: "age," "gender," "race/ethnicity," and tract-level census characteristics ("region," "population density," "urbanicity," "percent families below the poverty level," "percent housing units rented," "percent bachelor, graduate, or professional degrees," and "median household income").

For the "age" variable, three age groups were considered: 12 to 17, 18 to 25, and 26 or older. The latter group was the reference group in the models. "Gender" was entered in the models, with female as the reference level. The "race/ethnicity" variable had four levels—white, black or African American, other (non-Hispanic), and Hispanic or Latino—with white as the reference level.

The other variables were census tract-level characteristics. "Region" had four levels—Northeast, Midwest, South, and West—with West as the reference level. "Population density" had five levels, with "segment not in a core-based statistical area (CBSA) and in rural area" as the reference level. "Urbanicity" had four levels—rural, nonmetropolitan statistical area (MSA) urban, MSA < 1 million, and MSA  $\geq$  1 million—with rural as the reference level. The remaining variables were continuous variates: "percent families below poverty level," "percent housing units rented," "percent bachelor, graduate, or professional degrees," and "median household income (in multiples of \$10,000)." In addition, the models for response propensity for the second interview included two covariates that were based on first interview, reflecting having reported sensitive personal information (see Table 6.49). These covariates were "reported mental health problems" and "reported lifetime use of cocaine or heroin."

For comparison purposes, the response propensity models were run for the entire 2006 NSDUH sample.



### **6.3.1 First Interview Response Propensity Models**

The "age in years" of the individual selected in the sample showed high significance ( $p$  value = 0.00, Table 6.48). The younger the individual, the more likely he or she was to respond. Compared with the 26 or older age group (the reference group), 12 to 17 year olds were much more likely to respond to the first interview, with an odds ratio (OR) of 2.51. The 18 to 25 year olds were also more likely to respond than the 26 or older group (OR = 1.63), even though not as likely as the 12 to 17 group. "Race/ethnicity" also was significant ( $p$  value = 0.00), with Hispanics showing a significantly higher propensity to respond (OR = 1.77). "Gender" was not significant ( $p$  value = 0.76). None of the tract-level characteristics was significant ( $p$  values = 0.28 or higher).

### **6.3.2 Reinterview Response Propensity Models**

Similar to the findings for the first interview, the "age" of the individual selected in the sample showed high significance ( $p$  value = 0.00), with the odds declining as the age increased (Table 6.49). Compared with the 26 or older group, the 12 to 17 year olds were much more likely to respond to the reinterview (OR = 3.37), and the 18 to 25 year olds were also more likely to respond than the 26 or older group (OR = 1.60). Neither "gender" nor "race/ethnicity" was significant ( $p$  values = 0.35 and 0.18, respectively). Most of the tract-level characteristics were insignificant. However, two tract-level characteristics were highly significant: "percent housing units rented" ( $p$  value = 0.01) and "median household income" ( $p$  value = 0.00), with both having a negative effect on the propensity to respond to the reinterview.

### **6.3.3 Comparison with the Full 2006 NSDUH Sample**

For the full 2006 NSDUH sample, the "age" of the individual selected in the sample was significant ( $p$  value = 0.00) and showed the same pattern of the propensity to respond declining with age (Table 6.50). "Gender" also was significant ( $p$  value = 0.00), with males less likely to respond than females (OR = 0.82). In the case of the Reliability Study sample, males were slightly more likely to respond to the first interview (OR = 1.03), but less likely to respond to the reinterview (OR = 0.85), although not significantly so, possibly because of the smaller sample size.

"Race/ethnicity" was significant ( $p$  value = 0.00), with blacks (OR = 1.14) and Hispanics (OR = 1.13) more likely to respond than whites, and other non-Hispanics less likely to respond than whites (OR = 0.62). These differences in the race/ethnicity groups also were seen in the Reliability Study first interview response. With the exception of "region," none of the other tract-level characteristics had a significant effect on the response propensity in the full 2006 NSDUH sample. Within the regions, the Northeast had a lower propensity to respond (OR = 0.90) compared with the reference region (West), with highly nonsignificant differences between the West and Midwest regions or between the West and South regions.

## **6.4. Dependence of Age at First Use Reliability and Duration of Recall**

The reliability of age at first use under the exact agreement criterion was mostly moderate, with kappa values ranging from 0.40 to 0.59. See Section 4.1.2 and Landis and Koch

(1977) for categorization of the kappa. When the approximate agreement criterion was used, allowing for up to a 1-year difference, the reliability improved to Landis and Koch's "substantial" range (0.61 to 0.80 kappas) for the age at first use of alcohol, cigarettes, and marijuana and hashish. Respondents' ability to accurately recall their age at first use may decline as more years have passed since their first use. Therefore, logistic regression models were fitted to the data to explore this possibility, where the age at first use responses at the first interview (T1) and at the reinterview (T2) were compared.

Two levels of consistency between these age at first use responses were considered: exact agreement and approximate agreement, allowing for up to a 1-year difference. The dependent variable in the models was defined as 1 if the T1 and T2 age at first use responses agreed (exactly or approximately), and 0 otherwise. The covariate of interest was the number of years since first use. However, additional covariates that may have had an effect on the outcome were included in the models. These covariates were "gender," "race/ethnicity" (four levels), "education," and a covariate indicating whether both interviews were administered by the "same field interviewer" (FI). The current "age" (using T1 data) was not included in the models because of its high correlation with the number of years since first use, making it difficult to separate the effects of these two factors from each other. Indeed, the correlation between the number of years since first use and the current age was 0.97 for both cigarettes and for alcohol and 0.90 for marijuana and hashish. Because in the case of young individuals recall may be less of an issue, two age groups were modeled: 18 years old or older and 24 years old or older.

In this section, the length of time since first use is referred to as "duration of recall" (DOR) and is defined as the number of years since first use as reported at T1. Thus,  $DOR = AGE - AFU$ , where AGE is the current age and AFU is the age at first use, both calculated from T1 data.

#### **6.4.1 General Observations on the Consistency of Age at First Use of Cigarettes, Alcohol, and Marijuana and Hashish and the Years Since First Use**

Twelve models were fitted. These models were all combinations of the three substances (cigarettes, alcohol, and marijuana and hashish), two age groups (18 years old or older and 24 years old or older), and for exact agreement and approximate agreement (up to a 1-year difference). For the results, see Tables 6.51 to 6.57.

Eleven of the models showed the number of years since first use having a negative effect on the consistency between the two ages at first use, either exactly or approximately. In the case of approximate agreement for the marijuana and hashish age at first use, there was no effect. Although not all *p* values were significant, most were, and the uniformity in the direction of the effect of years since first use indicated that this phenomenon indeed exists.

#### **6.4.2 Effect of the Duration of Recall on the Consistency of Age at First Use of Cigarettes, Alcohol, and Marijuana and Hashish and the Years Since First Use**

Testing the effects of the duration of recall on exact agreement between the age at first use reported at T1 and that of T2 showed statistical significance. Similarly, when the agreement requirement was relaxed to allow up to a 1-year difference between the T1 and T2 age at first use

responses, the effects of the duration of recall were also significant. To determine the magnitude of these effects, predicted marginals were calculated using the model for the 18 or older age group for each of the three substances either for the outcome of exact agreement or with approximate agreement (allowing a 1-year difference). This model is referred to as "Model 1" in Table 6.57. Additionally, predicted marginals were calculated using the 12 or older group (i.e., all ages). "Education" was not included in this model because it is an inappropriate covariate in the case of those aged 12 to 17. The two models gave very close predicted margins.

### 6.4.3 Predicted Marginals and Their Interpretation

To help to appreciate the impact of the duration of recall on the agreement of the age at first use, predicted marginals are presented in Table 6.57. Predicted marginals at  $x = x_0$ , where  $x_0$  is a certain value of a covariate  $x$ , are defined as follows. Suppose every respondent had his or her covariates as in the data, except for the covariate  $x$ , which is set at  $x = x_0$ . Then the predicted marginal is the mean of the predicted values of these modified observations of the dependent variable. For example, in the hypothetical case where the duration of recall was set to 5 years for all those who had valid age at first use for cigarettes at both T1 and T2, Model 1 predicted 66 percent giving same age at first use at T1 and T2, and 85 percent having a difference of no more than 1 year between the age at first use at T1 and that at T2.

## 6.5. Hui-Walter Model and Findings

Suppose there is a latent dichotomous variable, such as the use or nonuse of a substance (in this case, past year use of marijuana). Let the true state of a dichotomous condition be denoted by  $X$ , where  $X = 1$  if the condition holds,  $X = 0$  otherwise. The responses given at the first interview (T1) and at the second interview (T2) are denoted by  $A$  and  $B$ , respectively, taking on the value of 1 if the response was positive and 0 otherwise. The false negative (FN) rate at T1 is therefore  $\Pr(A = 0|X = 1)$ . The false positive (FP) rate is  $\Pr(A = 1|X = 0)$ , and the rates at T2 are similarly represented, with  $B$  replacing  $A$ .

Although  $X$  is not directly observed, the true prevalence rate,  $\Pr(X = 1)$ , can be estimated, as well the reporting error rates, using the approach of Hui and Walter (1980). The Hui-Walter (H-W) approach assumes the population can be partitioned into subgroups defined by a grouping variable  $G$  where the following conditions hold:

1. The prevalence rates are different across the groups.
2. The FP and FN rates are the same across the groups.
3. The responses  $A$  and  $B$  are conditionally independent, given the true state of  $X$ .

In this study, "gender," denoted by  $S$  ( $S = 1$  for males), was used as the primary grouping variable, but additional grouping variables were introduced to study the variation of FP and FN rates by various demographic domains of interest, with  $G$  representing the collection of these additional grouping variables. To simplify the discussion, it was assumed without loss of generality that  $G$  was also dichotomous.

### 6.5.1 Hui-Walter Model

Let SGAB denote the observed  $2 \times 2 \times 2 \times 2$  table, and let  $\pi_{sgab} = \Pr(S = s, G = g, A = a, B = b)$ , the probability of an observation in cell (s,g,a,b). Further, let  $\pi_{sgab|x} = \Pr(S = s, G = g, A = a, B = b | x)$  and note that

$$\pi_{sgab} = \sum_x \pi_x \pi_{sgab|x}.$$

Under the H-W model, the above equation can be written as

$$\pi_{sgab} = \sum_x \pi_{sg} \pi_{x|sg} \pi_{a|xg} \pi_{b|xg}.$$

In this parameterization,  $\pi_{sg}$  is the proportion of the population belonging to domain (s,g). As an example, if  $G$  is "race/ethnicity" and  $G = 1$  denotes blacks, then  $\pi_{s=1,g=1}$  is the proportion of black males in the population. Then  $\pi_{x|sg}$  for  $s = 1, g = 1$ , is the prevalence of drug  $X$  for this subgroup;  $\pi_{a|xg}$  when  $a = 0, x = 1$ , and  $g = 1$  is the FN rate for blacks for the interview (T1); and  $\pi_{b|xg}$  is the corresponding rate for the reinterview (T2). The FP rate can be defined analogously when  $a = 1, x = 0$ , and  $g = 1$ . By varying the levels of  $G$ , the prevalence, FN, and FP rates can be defined for any group of interest defined by  $G$ .

The H-W model is identifiable for any variable  $G$ , and the prevalence and error rate parameters can be estimated using maximum likelihood methods. In the next section, the methodology is illustrated for past year marijuana use. In addition to "gender" and "race/ethnicity," other grouping variables used in this analysis included "age group," "U.S. born," and "metropolitan area."

### 6.5.2 Results

The software package  $\ell$ EM (Vermunt, 1997)<sup>7</sup> was used for producing the results in this section. The outcome variable was "past year marijuana use" (yes/no). Thus,  $X = 1$  denotes true past year marijuana use, and  $X = 0$  indicates no past year use. Likewise,  $A$  and  $B$  were defined analogously for the interview and reinterview observations, respectively. The following variables were included in the modeling:

- "past year marijuana use" (yes/no);
- "gender" (male/female);
- "race" (1 = white, non-Hispanic; 2 = black, non-Hispanic; 3 = other, non-Hispanic; 4 = Hispanic or Latino);
- "age group" (12 to 17, 18 to 25, 26 or older);
- "U.S. born" (yes/no); and

---

<sup>7</sup> For details, see <http://www.uvt.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html>.

- "metro" (1 = metropolitan, 2 = nonmetropolitan).

As previously noted, the H-W approach requires the assumption that the FP and FN rates be the same across the levels of at least one grouping variable. The plausibility of this assumption was tested for each of the five grouping variables by fitting a model that specified no interaction of the grouping variable FP and FN rates while the remaining four variables were allowed to interact. This approach allowed the hypothesis of no interaction to be tested for each grouping variable in turn. The results of this analysis failed to reject the required H-W assumption of no interaction of gender with the FP and FN rates. In fact, it was the only variable among the five tested that failed to exhibit significant interaction.

Tables 6.58 to 6.62 at the end of this chapter provide estimates of the true prevalence rates and the FN rates for NSDUH (i.e., indicator *A*) by each grouping variable in the model. All FP error rates were estimated to be 0 or very nearly 0 by the model across all grouping variables. The fitted model contained all five grouping variables simultaneously. Up to four-way interactions were included in the model; for example, the prevalence probabilities  $\pi_{x|sgh}$  were estimated for combinations of up to three grouping variables, *S*, *G*, and *H*. No constraints were imposed on the model other than those required by the H-W model noted earlier. All differences were significant at the .05 level of significance, and all analyses were conducted using weighted data tables. The analysis itself did not take the complex sampling design into consideration, however, so the results of statistical tests discussed in this section must be treated with some caution.

The FN rates were constrained to be equal for males and females (Table 6.58). However, as noted earlier, this assumption was tested and could not be rejected. Blacks and other non-Hispanics had very high error rates compared with whites and Hispanics (Table 6.59). This could be symptomatic of distrust of authority and/or the purposes of the survey. The model indicated much lower FN rates for those aged 18 to 25 than for the younger or the older age groups (Table 6.60). Perhaps this is an indication of greater perceived stigma among the very young and older persons. Past year users not U.S. born exhibited much higher rates of underreporting (Table 6.61). Again, this may be related to a fear of disclosure. The model also suggested somewhat higher FN rates for persons living in metropolitan areas than for those living in nonmetropolitan areas (Table 6.62).

These preliminary results indicate that males and females were equally likely to underreport past year marijuana use. Whites, Hispanics, and 18 to 25 year olds were less likely to underreport, while persons born outside the United States and persons living in metropolitan areas were more likely to underreport. More research is needed to determine the underlying causes of these variations in underreporting.

The preliminary results suggest that it is possible to obtain plausible estimates of the FP and FN rates using the H-W method. The model fits the data well, and its assumptions seem plausible. Moreover, assumptions that could be tested were verified. It is recommended that the H-W analysis be expanded to other drugs and substances in future NSDUH reinterview studies. The advantages of this method over reliability estimation are that it provides indicators of data quality that can be compared across demographic groups regardless of the within-group prevalence.

**Table 6.1. Reliability Statistics for Substance Use in Lifetime among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Substance Use Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Use (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Cigarettes</b>					
Daily Cigarette Use <sup>4</sup>	3,136	0.00	69.3	96.6 (0.7)	0.92 (0.02)
Smoked 100 Cigarettes	3,134	0.12	40.8	96.4 (0.8)	0.93 (0.02)
<b>Smokeless Tobacco<sup>5</sup></b>					
Snuff	3,121	0.44	45.7	97.6 (0.5)	0.95 (0.01)
Chewing Tobacco	3,135	0.00	17.6	96.3 (1.0)	0.88 (0.03)
	3,135	0.00	11.2	97.1 (0.6)	0.86 (0.03)
<b>Cigars</b>	3,134	0.00	13.7	96.0 (1.0)	0.84 (0.04)
<b>Pipe Tobacco</b>	3,133	0.04	37.3	94.5 (0.8)	0.88 (0.02)
	3,135	0.01	20.0	95.0 (0.9)	0.84 (0.03)
<b>Alcohol</b>	3,135	0.04	84.9	95.8 (0.8)	0.83 (0.03)
<b>Marijuana and Hashish</b>	3,132	0.07	37.6	96.8 (0.6)	0.93 (0.01)
<b>Cocaine</b>	3,132	0.17	13.3	98.3 (0.4)	0.92 (0.02)
Crack	3,132	0.17	3.2	99.0 (0.5)	0.82 (0.07)
<b>Heroin</b>	3,135	0.00	1.1	99.9 (0.0)	0.95 (0.02)
<b>Hallucinogens</b>	3,123	0.17	15.0	98.1 (0.4)	0.92 (0.02)
LSD	3,129	0.13	9.5	98.7 (0.3)	0.92 (0.02)
Ecstasy	3,133	0.04	5.2	99.3 (0.2)	0.93 (0.02)
<b>Inhalants</b>	3,125	0.22	8.0	97.9 (0.4)	0.84 (0.03)
<b>Pain Relievers</b>	3,129	0.08	12.4	95.4 (0.8)	0.78 (0.04)
<b>Tranquilizers</b>	3,127	0.09	7.2	97.0 (0.5)	0.78 (0.04)
<b>Stimulants</b>	3,130	0.10	7.9	96.7 (0.9)	0.77 (0.06)
Methamphetamine <sup>6</sup>	3,132	0.07	5.6	97.1 (0.9)	0.71 (0.08)
<b>Sedatives</b>	3,127	0.11	3.9	98.5 (0.4)	0.79 (0.06)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Daily Cigarette Use is defined as ever smoking every day for at least 30 days.

<sup>5</sup> Smokeless tobacco is defined as the use of snuff or chewing tobacco.

<sup>6</sup> This methamphetamine estimate is based only on questions from the core stimulants module and does not include information from the noncore special drugs module.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.2. Reliability Statistics for Substance Use in the Past Year among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Substance Use Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Use (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Cigarettes</b>	3,129	0.19	28.5	97.1 (0.6)	0.93 (0.01)
<b>Smokeless Tobacco<sup>4</sup></b>	3,133	0.10	3.4	99.0 (0.3)	0.84 (0.04)
Snuff	3,134	0.10	2.7	99.2 (0.2)	0.84 (0.04)
Chewing Tobacco	3,133	0.00	1.6	99.0 (0.3)	0.68 (0.08)
<b>Cigars</b>	3,131	0.28	8.9	95.4 (0.9)	0.72 (0.05)
<b>Alcohol</b>	3,130	0.09	65.9	95.4 (0.9)	0.90 (0.02)
<b>Marijuana and Hashish</b>	3,128	0.11	9.8	96.9 (0.7)	0.82 (0.03)
<b>Cocaine</b>	3,132	0.17	2.5	99.0 (0.4)	0.76 (0.07)
<b>Hallucinogens</b>	3,119	0.21	1.4	99.2 (0.2)	0.72 (0.07)
LSD	3,127	0.17	0.2	99.8 (0.1)	* (*)
Ecstasy	3,131	0.08	0.9	99.4 (0.2)	* (*)
<b>Inhalants</b>	3,120	0.23	0.9	99.6 (0.2)	* (*)
<b>Pain Relievers</b>	3,123	0.37	4.8	97.8 (0.4)	0.73 (0.05)
<b>Tranquilizers</b>	3,125	0.10	1.9	99.1 (0.4)	0.72 (0.09)
<b>Stimulants</b>	3,127	0.13	1.1	99.6 (0.1)	* (*)
Methamphetamine-Composite <sup>5</sup>	3,132	0.04	0.4	99.5 (0.2)	* (*)
Methamphetamine-Core <sup>5</sup>	3,130	0.11	0.3	99.6 (0.2)	* (*)
Methamphetamine-Noncore <sup>5</sup>	2,972	0.00	0.0	100.0 (0.0)	* (*)
<b>Sedatives</b>	3,125	0.13	0.6	99.2 (0.4)	* (*)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Respondents who reported no lifetime use of a substance were assigned a 'no' value to the corresponding past year variable even though the question was not asked.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Smokeless tobacco is defined as the use of snuff or chewing tobacco.

<sup>5</sup> The core methamphetamine estimate is based only on questions from the core stimulants module and does not include information from the noncore special drugs module. The noncore methamphetamine estimate is based on questions from the noncore special drugs module, but the denominator for the calculation of "percent reporting consistently" consists of those who responded "never used" or "don't know" to the lifetime methamphetamine use question in the core stimulants module. The composite methamphetamine estimate incorporates responses from the core stimulants module and information from the noncore special drugs module for respondents who indicated that they did not originally report methamphetamine use because they did not think of methamphetamine as a prescription drug.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.3. Reliability Statistics for Age at First Use of Substances among Lifetime Users Aged 12 or Older: 2006 NSDUH Reliability Study**

Age at First Substance Use Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
<b>Cigarettes</b>	1,751	1.26	61.8 (3.0)	0.58 (0.03)	77.5 (3.1)	0.71 (0.04)	0.11 (0.03)
Daily Cigarette Use <sup>5</sup>	917	0.97	48.4 (3.9)	0.43 (0.04)	73.5 (3.8)	0.64 (0.05)	0.12 (0.03)
<b>Snuff</b>	323	0.33	54.5 (6.7)	0.51 (0.07)	74.0 (5.6)	0.68 (0.07)	0.11 (0.05)
<b>Chewing Tobacco</b>	296	0.96	39.1 (6.5)	0.35 (0.07)	59.0 (7.6)	0.52 (0.08)	0.15 (0.07)
<b>Cigars</b>	905	2.25	39.4 (3.7)	0.34 (0.04)	58.8 (4.2)	0.50 (0.05)	0.11 (0.02)
<b>Alcohol</b>	2,181	1.69	55.8 (2.6)	0.51 (0.03)	75.8 (2.5)	0.68 (0.03)	0.39 (0.11)
<b>Marijuana and Hashish</b>	1,160	1.67	53.8 (3.6)	0.50 (0.04)	80.1 (3.2)	0.74 (0.04)	0.07 (0.02)
<b>Cocaine</b>	351	3.21	49.2 (5.8)	0.45 (0.06)	70.3 (5.5)	0.63 (0.07)	0.19 (0.05)
<b>Hallucinogens</b>	430	0.91	57.1 (5.6)	0.53 (0.06)	86.8 (3.3)	0.83 (0.04)	0.05 (0.02)
<b>LSD<sup>6</sup></b>	234	2.38	58.8 (6.5)	0.54 (0.07)	89.2 (2.7)	0.85 (0.04)	0.08 (0.05)
<b>Ecstasy<sup>7</sup></b>	192	2.62	51.1 (7.9)	0.47 (0.09)	83.2 (5.0)	0.79 (0.06)	0.05 (0.02)
<b>Inhalants</b>	245	1.00	47.2 (6.7)	0.42 (0.07)	67.1 (5.6)	0.57 (0.07)	0.15 (0.06)
<b>Pain Relievers</b>	365	3.86	32.6 (4.5)	0.29 (0.05)	55.6 (6.5)	0.49 (0.07)	0.09 (0.02)
<b>Tranquilizers</b>	197	4.82	31.6 (7.7)	0.27 (0.08)	77.6 (6.0)	0.73 (0.07)	0.12 (0.09)
<b>Stimulants<sup>8</sup></b>	178	0.33	49.7 (6.9)	0.45 (0.07)	74.4 (7.1)	0.67 (0.08)	0.12 (0.04)
Methamphetamine <sup>9</sup>	82	0.00	48.5 (10.3)	0.42 (0.11)	77.3 (10.2)	0.69 (0.15)	0.07 (0.05)
<b>Sedatives</b>	52	0.57	23.5 (12.1)	0.21 (0.10)	52.1 (17.2)	0.47 (0.16)	0.15 (0.06)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Respondents were instructed to provide their age at first use, in whole years, for each substance ever used. All variables in this table are among lifetime users of the respective substances. The agreement criterion for approximate agreement is relaxed by 1 year of age.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> Daily Cigarette Use is defined as ever smoking every day for at least 30 days.

<sup>6</sup> This estimate was created based on two questions: One was asked of respondents who reported using LSD only, and the other of respondents who reported using LSD plus another hallucinogen.

<sup>7</sup> This estimate was created based on two questions: One was asked of respondents who reported using Ecstasy only, and the other of respondents who reported using Ecstasy plus another hallucinogen.

<sup>8</sup> This estimate was created as the minimum value of the age at last use of methamphetamine and of other stimulants.

<sup>9</sup> Age at first use of methamphetamine was asked about in the core module; thus, among lifetime users of methamphetamine, it was based on questions from only the core stimulants module.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).



**Table 6.4. Reliability Statistics for First Substance Used among Lifetime Users Aged 12 or Older of Alcohol, Marijuana, or Cigarettes Who Reported the Same Age at First Use for at Least Two of the Substances: 2006 NSDUH Reliability Study**

First Substance Used Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
Alcohol or Cigarettes <sup>3</sup>	136	1.11	70.9 (10.2)	0.43 (0.19)
Marijuana or Cigarettes <sup>4</sup>	69	0.00	91.0 (3.1)	0.81 (0.07)
Alcohol or Marijuana <sup>5</sup>	81	0.00	83.1 (8.0)	0.26 (0.16)
Alcohol, Cigarettes, or Marijuana <sup>6</sup>	63	3.23	91.8 (4.3)	0.84 (0.08)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Response options for this question are "alcohol" and "cigarettes." This question was asked of respondents who had the same age at first use for alcohol and cigarettes, but not for marijuana.

<sup>4</sup> Response options for this question are "marijuana" and "cigarettes." This question was asked of respondents who had the same age at first use for marijuana and cigarettes, but not for alcohol.

<sup>5</sup> Response options for this question are "alcohol; and marijuana." This question was asked of respondents who had the same age at first use for alcohol and marijuana, but not for cigarettes.

<sup>6</sup> Response options for this question are "alcohol," "cigarettes," and "marijuana." This question was asked of respondents who had the same age at first use for alcohol, cigarettes, and marijuana.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.5. Reliability Statistics for Measures of Alcohol, Marijuana, and Cocaine Use among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
Number of Months from First Alcohol Use to the Initial Interview <sup>5</sup>	121	47.75	68.1 (6.3)	0.65 (0.07)	74.8 (5.7)	0.71 (0.07)	0.36 (0.15)
Number of Days Used in the Past Year <sup>6</sup>							
Alcohol	1,807	1.31	30.9 (2.8)	0.28 (0.03)	45.8 (2.9)	0.41 (0.03)	0.18 (0.03)
Marijuana	445	1.75	25.8 (3.6)	0.23 (0.04)	46.6 (4.8)	0.42 (0.05)	0.15 (0.04)
Cocaine	90	0.00	24.2 (7.7)	0.22 (0.07)	51.4 (9.7)	0.47 (0.09)	0.52 (0.18)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> This estimate was created using month and year of first use, which were asked only of respondents who reported initiating alcohol use at their current age or 1 year younger than their current age. Number of months from first alcohol use to the initial interview was calculated as the number of months from the month of initiation to the month of the initial interview. The agreement criterion for approximate agreement was relaxed by 1 month.

<sup>6</sup> Respondents were asked to report the number of days used in the past 12 months, the number of days used in each month during the past 12 months on average, or the number of days used each week during the past 12 months on average, depending on which method they had indicated was easiest. These questions were asked only of respondents who were past year users of the respective substance. The agreement criterion for approximate agreement was relaxed by 5 days.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.6. Reliability Statistics for Blunt Use among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Blunt Use Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Use (Weighted) <sup>3</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
				Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
Lifetime Use	3,133	0.02	14.0	96.4 (0.6)	0.84 (0.03)	N/A	N/A	N/A
Past Year Use <sup>6</sup>	3,128	0.04	5.1	97.1 (0.6)	0.66 (0.05)	N/A	N/A	N/A
Age at First Use <sup>7</sup>	573	1.31	N/A	38.2 (4.1)	0.32 (0.04)	71.8 (4.6)	0.62 (0.06)	0.17 (0.08)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

N/A: Not applicable.

NOTE: Blunts are defined as cigars with marijuana in them.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.  
<sup>5</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>6</sup> This question was asked only of respondents who previously indicated ever using blunts. Respondents who reported no lifetime use of blunts were assigned a "no" value to the past year variable even though the question was not asked.

<sup>7</sup> Respondents were instructed to provide their age at first use in whole years. This question was asked only of respondents who previously reported ever using blunts. The agreement criterion for approximate agreement was relaxed by 1 year of age.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.7. Reliability Statistics for Lifetime Nonmedical Use of Specific Noncore Stimulants, Sedatives, Cough and Cold Medicines, or Hallucinogens and Past Year Nonmedical Use of Adderall® or Cough and Cold Medicines among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Substance Use Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Use (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Lifetime Use</b>					
Methamphetamine <sup>4</sup>	2,971	0.03	1.2	99.4 (0.2)	* (*)
GHB	3,129	0.04	0.9	99.8 (0.1)	* (*)
Nonmedical Use of Adderall®	3,132	0.01	2.0	99.4 (0.1)	0.85 (0.04)
Nonmedical Use of Ambien®	3,136	0.00	1.0	99.3 (0.2)	0.65 (0.09)
Nonprescription Cough or Cold Medicine	3,135	0.05	2.9	97.1 (0.9)	0.49 (0.10)
Ketamine	3,132	0.35	1.6	99.5 (0.2)	0.81 (0.10)
DMT, AMT, or 5-MeO-DIPT	3,131	0.10	0.3	99.8 (0.1)	* (*)
<i>Salvia Divinorum</i>	3,131	0.38	0.6	99.8 (0.1)	* (*)
Needle Use <sup>5</sup>	3,126	0.23	1.4	99.2 (0.3)	0.74 (0.11)
<b>Past Year Use<sup>6</sup></b>					
Nonmedical Use of Adderall®	3,131	0.01	0.7	99.6 (0.1)	* (*)
Nonprescription Cough or Cold Medicine	3,131	0.08	0.5	99.5 (0.2)	* (*)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

AMT = alpha-methyltryptamine; DMT = dimethyltryptamine; 5-MeO-DIPT = 5-methoxy-diisopropyltryptamine ("Foxy"); GHB = gamma hydroxybutyrate.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> The methamphetamine estimate is based strictly on questions from the noncore special drugs module that were asked only of persons who responded "never used" or "don't know" to the lifetime methamphetamine use question in the core stimulants module.

<sup>5</sup> Refers to lifetime use of a needle to inject a drug that was not prescribed or that was taken only for the experience or feeling it caused.

<sup>6</sup> Respondents who reported no lifetime use of a substance were assigned a "no" value to the corresponding past year variable even though the question was not asked.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.8. Reliability Statistics for Perceived Risk and Availability of Substances, Participation in Risky Behaviors, and Likelihood of Smoking Cigarettes for Nonsmokers among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	All Response Options (Categorical) <sup>3</sup>		Dichotomized Response <sup>3</sup>		
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Percent Reporting Factor (Weighted) <sup>4</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Perceived Risk<sup>5</sup></b>							
Smoking 1 or More Packs of Cigarettes Per Day	3,108	1.18	77.7 (1.6)	0.46 (0.04)	75.4	80.7 (1.5)	0.51 (0.04)
Smoking Marijuana Once a Month	3,094	1.87	62.7 (2.2)	0.48 (0.03)	37.6	81.8 (1.7)	0.60 (0.04)
Smoking Marijuana Once or Twice a Week	3,097	2.05	69.0 (2.0)	0.51 (0.03)	51.2	81.8 (1.7)	0.64 (0.03)
Trying LSD Once or Twice	3,010	2.87	76.7 (1.6)	0.51 (0.03)	68.8	84.4 (1.6)	0.63 (0.04)
Using LSD Once or Twice a Week	3,025	2.27	86.7 (1.4)	0.35 (0.05)	88.4	87.8 (1.3)	0.39 (0.05)
Trying Heroin Once or Twice	3,070	1.88	80.9 (1.5)	0.45 (0.04)	79.7	83.9 (1.5)	0.51 (0.04)
Using Heroin Once or Twice a Week	3,081	1.58	92.7 (1.3)	0.44 (0.09)	93.1	93.1 (1.3)	0.46 (0.09)
Using Cocaine Once a Month	3,071	1.76	75.2 (1.9)	0.49 (0.03)	68.1	84.3 (1.4)	0.65 (0.03)
Using Cocaine Once or Twice a Week	3,081	1.42	88.2 (1.4)	0.40 (0.06)	88.5	90.3 (1.4)	0.50 (0.07)
Having 4 or 5 Drinks of Alcohol Nearly Every Day	3,117	0.65	73.1 (1.8)	0.42 (0.04)	67.4	79.0 (1.8)	0.51 (0.04)
Having 5 or More Drinks of Alcohol Once or Twice a Week	3,105	1.03	57.0 (2.0)	0.33 (0.03)	39.6	75.9 (1.5)	0.50 (0.03)
<b>Perceived Availability<sup>6</sup></b>							
Marijuana	3,000	5.93	62.5 (2.3)	0.52 (0.03)	56.5	87.4 (1.5)	0.74 (0.03)
LSD	2,885	8.30	61.6 (2.3)	0.48 (0.03)	18.4	87.6 (1.4)	0.55 (0.04)
Cocaine	2,947	7.30	62.2 (2.3)	0.51 (0.03)	28.7	88.1 (1.3)	0.71 (0.03)
Crack	2,941	7.05	59.5 (2.2)	0.48 (0.03)	27.6	89.6 (1.1)	0.73 (0.03)
Heroin	2,924	7.62	59.9 (2.4)	0.45 (0.03)	16.5	91.0 (1.1)	0.67 (0.04)

(continued)

**Table 6.8. Reliability Statistics for Perceived Risk and Availability of Substances, Participation in Risky Behaviors, and Likelihood of Smoking Cigarettes for Nonsmokers among Persons Aged 12 or Older: 2006 NSDUH Reliability Study (continued)**

Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	All Response Options (Categorical) <sup>3</sup>		Dichotomized Response <sup>3</sup>		
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Percent Reporting Factor (Weighted) <sup>4</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Participation in Risky Behavior<sup>7</sup></b>							
Get a Real Kick Out of Doing Dangerous Things	3,118	0.30	72.8 (2.1)	0.58 (0.03)	2.1	97.8 (0.4)	0.43 (0.07)
Like to Test Yourself by Doing Risky Things	3,124	0.16	74.4 (1.8)	0.60 (0.03)	1.7	98.4 (0.4)	0.50 (0.08)
Wear a Seatbelt When You Ride in Front Passenger Seat of Car	3,135	0.03	91.3 (1.1)	0.72 (0.03)	82.8	94.7 (1.0)	0.82 (0.03)
Wear a Seatbelt When You Drive a Car	3,128	0.07	91.2 (1.2)	0.78 (0.03)	76.8	93.8 (1.0)	0.84 (0.02)
<b>Likelihood of Smoking Cigarettes<sup>8</sup></b>							
Would Smoke Cigarettes if Best Friend Offered	711	0.08	87.5 (2.4)	0.53 (0.07)	1.9	95.8 (1.7)	0.21 (0.11)
Will Smoke Cigarettes in the Next 12 Months	711	0.30	77.7 (1.6)	0.46 (0.04)	0.6	98.3 (0.8)	0.23 (0.14)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> The response options for both the All Response Options (Categorical) and the Dichotomized Response columns are specified in footnotes for each of the individual row headers.

<sup>4</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>5</sup> Response options for these questions are "no risk," "slight risk," "moderate risk," and "great risk." The dichotomized response corresponds to persons reporting "great risk."

<sup>6</sup> Response options for these questions are "probably impossible," "very difficult," "fairly difficult," "fairly easy," and "very easy." The dichotomized response corresponds to persons reporting "fairly easy" or "very easy."

<sup>7</sup> Response options for these questions are "never," "seldom," "sometimes," and "always." The dichotomized response corresponds to persons reporting "always."

<sup>8</sup> Response options for these questions are "definitely yes," "probably yes," "probably not," and "definitely not." The dichotomized response corresponds to persons reporting "definitely yes" or "probably yes." These questions were asked only of respondents aged 12 to 17 who had never used cigarettes.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.9. Reliability Statistics for Alcohol Dependence or Abuse in the Past Year and the Criteria of Alcohol Dependence or Abuse among Past Year Alcohol Users Aged 12 or Older: 2006 NSDUH Reliability Study**

Dependence or Abuse and Criteria Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Condition/Symptom (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Dependence<sup>4</sup></b>					
Need More to Get Same Effect/Same Amount Causes Less Effect	1,445	1.65	9.1	96.0 (0.9)	0.75 (0.07)
Reported Withdrawal Symptoms	1,456	0.86	16.1	90.1 (1.5)	0.61 (0.07)
Set Limits But Not Able to Keep to Them	1,459	0.48	4.2	96.2 (2.0)	0.36 (0.16)
Wanted to Cut or Stop Using, But Not Able Month or More When Spent a Lot of Time Getting, Using, or Getting over Effects of Alcohol	1,465	0.01	7.5	95.6 (1.0)	0.68 (0.11)
Spent Less Time Doing Hobbies/Activities Due to Use of Alcohol	1,457	0.66	5.8	93.7 (1.1)	0.48 (0.16)
Continued Using Despite Emotional or Physical Problems It Caused	1,465	0.06	19.8	88.2 (1.9)	0.60 (0.06)
<b>Abuse<sup>5</sup></b>					
Serious Problems at Home, Work, or School Regularly Use and Do Something to Put Self in Danger	1,465	0.02	4.4	96.7 (0.6)	0.63 (0.08)
Use Caused Respondent to Do Things That Repeatedly Caused Trouble with the Law	1,465	0.29	6.6	96.6 (0.6)	0.72 (0.10)
Continued Using Despite Problems with Family/Friends	1,464	0.02	12.4	91.6 (1.3)	0.62 (0.07)
	1,465	0.02	3.3	95.8 (1.9)	0.44 (0.14)
	1,465	0.02	9.3	92.5 (1.3)	0.58 (0.09)
	1,465	0.02	1.6	99.0 (0.3)	* (*)
	1,463	0.03	6.4	94.2 (2.0)	0.45 (0.12)
<b>Dependence or Abuse</b>	1,446	1.63	15.2	90.9 (1.5)	0.64 (0.06)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Dependence or abuse is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. All variables in this table are among past year users of alcohol. Estimates here differ from estimates in NSDUH's 12-month analysis tables, which are among the total population.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Dependence on alcohol is defined as meeting at least three out of the seven dependence criteria included in the DSM-IV.

<sup>5</sup> Abuse of alcohol is defined as meeting at least one out of the four abuse criteria. Unlike the standard definition, it is defined independently from dependence, meaning that a respondent can be classified as having dependence and as having abused.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.10. Reliability Statistics for Marijuana Dependence or Abuse in the Past Year and the Criteria of Marijuana Dependence or Abuse among Past Year Marijuana Users Aged 12 or Older: 2006 NSDUH Reliability Study**

Dependence or Abuse and Criteria Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Condition/Symptom (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Dependence<sup>4</sup></b>					
Need More to Get Same Effect/Same Amount Causes Less Effect	312	2.01	15.4	84.6 (5.6)	0.53 (0.15)
Set Limits But Not Able to Keep to Them	313	0.00	29.8	87.9 (2.5)	0.72 (0.06)
Wanted to Cut or Stop Using, But Not Able Month or More When Spent a Lot of Time Getting, Using, or Getting over Effects of Marijuana	313	0.00	12.7	90.2 (4.2)	0.60 (0.19)
Spent Less Time Doing Hobbies/Activities Due to Use of Marijuana	313	0.00	7.0	88.3 (5.3)	0.32 (0.19)
Continued Using Despite Emotional or Physical Problems It Caused	312	2.01	48.8	73.9 (5.3)	0.48 (0.11)
<b>Abuse<sup>5</sup></b>					
Serious Problems at Home, Work, or School Regularly Use and Do Something to Put Self in Danger	313	0.00	18.6	88.3 (3.1)	0.61 (0.11)
Use Caused Respondent to Do Things That Repeatedly Caused Trouble with the Law	313	0.00	14.1	90.8 (4.1)	0.61 (0.17)
Continued Using Despite Problems with Family/Friends	313	0.00	16.4	86.6 (4.4)	0.55 (0.13)
	313	0.00	9.7	93.9 (2.1)	0.66 (0.12)
	313	0.00	9.4	93.2 (2.4)	0.64 (0.13)
	313	0.00	2.6	94.6 (3.9)	0.29 (0.25)
	313	0.00	5.0	96.3 (1.5)	0.65 (0.13)
	312	2.01	25.3	84.2 (5.4)	0.63 (0.12)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Dependence or abuse is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). All variables in this table are among past year marijuana users. Estimates here differ from estimates in NSDUH's 12-month analysis tables, which are among the total population.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Dependence on marijuana is defined as meeting at least three out of the six dependence criteria included in the DSM-IV.

<sup>5</sup> Abuse of marijuana is defined as meeting at least one out of the four abuse criteria. Unlike the standard definition, it is defined independently from dependence, meaning that a respondent can be classified as having dependence and as having abused.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).



**Table 6.11. Reliability Statistics for Cocaine Dependence or Abuse in the Past Year and the Criteria of Cocaine Dependence or Abuse among Past Year Cocaine Users Aged 12 or Older: 2006 NSDUH Reliability Study**

<b>Dependence or Abuse and Criteria Variable</b>	<b>Number of Respondents<sup>1</sup></b>	<b>Weighted Percent Missing<sup>2</sup></b>	<b>Percent Reporting Condition/Symptom (Weighted)<sup>3</sup></b>	<b>Weighted Percent Reporting Consistently (SE)</b>	<b>Kappa (SE)</b>
<b>Dependence<sup>4</sup></b>					
Need More to Get Same Effect/Same Amount Causes Less Effect	89	7.12	18.9	81.8 (7.5)	0.54 (0.19)
Reported Withdrawal Symptoms	90	0.37	25.9	84.3 (7.9)	0.65 (0.17)
Set Limits But Not Able to Keep to Them	91	0.21	14.9	86.2 (6.4)	0.50 (0.23)
Wanted to Cut or Stop Using, But Not Able Month or More When Spent a Lot of Time Getting, Using, or Getting over Effects of Cocaine	92	0.00	11.9	82.8 (7.2)	0.37 (0.19)
Spent Less Time Doing Hobbies/Activities Due to Use of Cocaine	91	0.21	10.0	80.0 (7.8)	0.09 (0.14)
Continued Using Despite Emotional or Physical Problems It Caused	92	0.00	32.5	87.9 (4.7)	0.76 (0.10)
	91	0.21	21.8	73.4 (9.5)	0.29 (0.17)
<b>Abuse<sup>5</sup></b>					
Serious Problems at Home, Work, or School Regularly Use and Do Something to Put Self in Danger	90	6.96	16.0	78.0 (8.7)	0.34 (0.20)
Use Caused Respondent to Do Things That Repeatedly Caused Trouble with the Law	90	6.96	18.1	93.3 (3.0)	0.84 (0.08)
Continued Using Despite Problems with Family/Friends	91	0.21	13.6	96.3 (2.3)	0.89 (0.07)
	89	7.57	15.5	89.0 (6.8)	0.66 (0.19)
<b>Dependence or Abuse</b>					
	91	0.21	9.4	99.6 (0.3)	0.98 (0.01)
	90	6.96	8.4	96.8 (1.6)	0.86 (0.08)
	89	7.12	24.8	84.2 (7.0)	0.65 (0.15)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Dependence or abuse is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). All variables in this table are among past year users of cocaine. Estimates here differ from estimates in NSDUH's 12-month analysis tables, which are among the total population.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Dependence on cocaine is defined as meeting at least three out of the seven dependence criteria included in the DSM-IV.

<sup>5</sup> Abuse of cocaine is defined as meeting at least one out of the four abuse criteria. Unlike the standard definition, it is defined independently from dependence, meaning that a respondent can be classified as having dependence and as having abused.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.12. Reliability Statistics for Pain Reliever Dependence or Abuse in the Past Year and the Criteria of Pain Reliever Dependence or Abuse among Past Year Pain Reliever Users Aged 12 or Older: 2006 NSDUH Reliability Study**

<b>Dependence or Abuse and Criteria Variable</b>	<b>Number of Respondents<sup>1</sup></b>	<b>Weighted Percent Missing<sup>2</sup></b>	<b>Percent Reporting Condition/Symptom (Weighted)<sup>3</sup></b>	<b>Weighted Percent Reporting Consistently (SE)</b>	<b>Kappa (SE)</b>
<b>Dependence<sup>4</sup></b>					
Need More to Get Same Effect/Same Amount Causes Less Effect	133	3.46	4.3	97.4 (1.9)	0.61 (0.21)
Reported Withdrawal Symptoms	136	0.24	10.4	89.5 (4.6)	0.37 (0.14)
Set Limits But Not Able to Keep to Them	137	0.04	6.7	89.7 (4.9)	0.15 (0.13)
Wanted to Cut or Stop Using, But Not Able	136	0.38	4.0	96.8 (3.1)	0.21 (0.20)
Month or More When Spent a Lot of Time Getting, Using, or Getting over Effects of Pain Relievers	136	0.13	0.5	99.9 (0.1)	* (*)
Spent Less Time Doing Hobbies/Activities Due to Use of Pain Relievers	137	3.01	16.3	90.2 (4.5)	0.52 (0.18)
Continued Using Despite Emotional or Physical Problems It Caused	138	0.03	2.4	97.5 (1.8)	0.39 (0.25)
<b>Abuse<sup>5</sup></b>					
Serious Problems at Home, Work, or School	138	0.03	3.7	97.6 (1.8)	0.58 (0.22)
Regularly Use and Do Something to Put Self in Danger	138	0.03	6.7	94.6 (3.3)	0.23 (0.18)
Use Caused Respondent to Do Things That Repeatedly Caused Trouble with the Law	138	0.03	4.0	99.8 (0.1)	* (*)
Continued Using Despite Problems with Family/Friends	138	0.03	4.3	94.1 (3.3)	0.06 (0.09)
<b>Dependence or Abuse</b>					
	138	0.03	1.5	99.9 (0.1)	* (*)
	138	0.03	1.5	99.1 (0.6)	* (*)
	133	3.46	8.5	93.0 (3.8)	0.34 (0.18)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Dependence or abuse is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). All variables in this table are among past year pain reliever users. Estimates here differ from estimates in NSDUH's 12-month analysis tables, which are among the total population.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Dependence on pain relievers is defined as meeting at least three out of the seven dependence criteria included in the DSM-IV.

<sup>5</sup> Abuse of pain relievers is defined as meeting at least one out of the four abuse criteria. Unlike the standard definition, it is defined independently from dependence, meaning that a respondent can be classified as having dependence and as having abused.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.13. Reliability Statistics for Substance Dependence or Abuse in the Past Year among Respective Substance Users Aged 12 or Older: 2006 NSDUH Reliability Study**

Substance Dependence or Abuse Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Condition (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Past Year Dependence<sup>4</sup></b>					
Illicit Drugs	441	3.67	18.6	88.7 (3.8)	0.68 (0.10)
Illicit Drugs or Alcohol	1,519	1.91	10.3	95.7 (0.9)	0.77 (0.06)
Illicit Drugs and Alcohol	368	2.33	12.5	90.5 (3.7)	0.56 (0.16)
<b>Past Year Abuse<sup>5</sup></b>					
Illicit Drugs	451	1.69	21.2	87.8 (3.1)	0.65 (0.08)
Illicit Drugs or Alcohol	1,550	0.24	14.0	91.3 (1.4)	0.65 (0.06)
Illicit Drugs and Alcohol	369	0.46	12.5	91.7 (2.4)	0.64 (0.11)
<b>Past Year Dependence or Abuse<sup>4,5</sup></b>					
Illicit Drugs	442	3.57	28.1	83.8 (3.9)	0.62 (0.09)
Illicit Drugs or Alcohol	1,522	1.84	17.0	90.6 (1.4)	0.67 (0.05)
Illicit Drugs and Alcohol	367	2.63	19.0	87.4 (3.9)	0.58 (0.12)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Dependence or abuse of illicit drugs or alcohol is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Respondents were classified as being dependent on illicit drugs if they were classified as being dependent on any of the specific illicit drugs, including, but not limited to, marijuana, cocaine, and pain relievers. These estimates are among past year users. Estimates here differ from estimates in NSDUH's 12-month analysis tables, which are among the total population.

<sup>5</sup> Respondents were classified as having abused illicit drugs if they were classified as abusing any of the specific illicit drugs, including, but not limited to, marijuana, cocaine, and pain relievers. Unlike the standard definition, it is defined independently from dependence, meaning that a respondent can be classified as having dependence and as having abused. These estimates are among past year users. Estimates here differ from estimates in NSDUH's 12-month analysis tables, which are among the total population.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.14. Reliability Statistics for Encounters with the Police or the Court System, Driving under the Influence (DUI) of Alcohol or Illicit Drugs in the Past Year, and Legal Knowledge about Marijuana among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Item (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Arrested and Booked</b>					
Breaking Law in Lifetime	3,126	0.26	16.2	92.3 (1.3)	0.70 (0.05)
Breaking Law in the Past Year <sup>4</sup>	3,125	0.26	2.9	98.1 (0.5)	0.66 (0.07)
DUI in the Past Year <sup>5</sup>	3,126	0.26	0.7	99.7 (0.1)	* (*)
Possession, Manufacture, or Sale of Drugs in the Past Year <sup>5</sup>	3,126	0.26	0.8	99.3 (0.3)	* (*)
<b>On Probation in the Past Year</b>	3,130	0.04	1.8	98.9 (0.3)	0.70 (0.08)
<b>On Parole/Supervised Release in the Past Year</b>	3,130	0.04	0.8	99.4 (0.2)	* (*)
<b>DUI in the Past Year<sup>6</sup></b>					
Alcohol	3,134	0.03	11.2	95.5 (0.9)	0.78 (0.04)
Illegal Drugs	3,134	0.03	3.4	97.9 (0.6)	0.69 (0.06)
<b>Marijuana Has Been Legally Approved in State for Medical Use</b>	2,299	24.07	20.0	93.4 (1.2)	0.79 (0.04)
<b>No Legal Penalty in State for First Possession of ≤ 1 Ounce of Marijuana<sup>7</sup></b>	2,041	37.97	3.7	94.9 (1.5)	0.07 (0.04)
<b>Sure of Maximum Legal Penalty<sup>8</sup></b>	2,028	0.48	46.3	77.4 (2.6)	0.55 (0.05)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Respondents who reported not being arrested and booked in lifetime were assigned a "no" value to the corresponding past year variable even though the question was not asked. This estimate is based on a write-in response to the number of times arrested and booked in the past year.

<sup>5</sup> Respondents who reported not being arrested and booked in lifetime or past year were assigned a "no" value for this variable even though the question was not asked.

<sup>6</sup> These questions were asked only among users of the respective substances. Respondents who were nonusers were assigned a value of "no" for that substance.

<sup>7</sup> Response options for this question are "a fine," "probation," "community service," "a possible prison sentence," "a mandatory prison sentence," and "no penalty." This estimate corresponds to persons reporting "no penalty."

<sup>8</sup> Response options for this question are "very sure," "somewhat sure," and "not at all sure." This estimate corresponds to persons reporting "very sure" or "somewhat sure." This question was asked only of respondents who provided a valid response to the previous question about the maximum legal penalty for possession of marijuana for own use.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.15. Reliability Statistics for Marijuana Purchase and Acquisition among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

<b>Purchase and Acquisition Variable</b>	<b>Number of Respondents<sup>1</sup></b>	<b>Weighted Percent Missing<sup>2</sup></b>	<b>Percent Reporting Activity (Weighted)<sup>3</sup></b>	<b>Weighted Percent Reporting Consistently (SE)</b>	<b>Kappa (SE)</b>	<b>Index of Inconsistency (SE)<sup>4</sup></b>
Method of Obtaining Most Recently Used Marijuana among Past Year Users <sup>5</sup>	441	3.02	N/A	90.4 (2.2)	0.81 (0.04)	0.18 (0.04)
Method of Obtaining Most Recently Used Marijuana among Persons Who Used within the Past Year But Not in the Past Month <sup>5,6</sup>	143	0.00	N/A	94.0 (3.1)	0.87 (0.07)	0.13 (0.07)
Last Bought Any Marijuana in the Past Year among Past Year Users <sup>7</sup>	445	1.69	56.5	90.2 (3.4)	0.80 (0.07)	N/A
Last Bought Any Marijuana in the Past Year, But Not in the Past Month, among Past Year Users <sup>7,8</sup>	277	2.45	41.8	85.9 (4.9)	0.69 (0.11)	N/A

\* Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

N/A: Not applicable.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> Response options for this question are "you bought it," "you traded something else for it," "you got it for free or shared someone else's," and "you grew it yourself." This question was asked only among past year marijuana users.

<sup>6</sup> Respondents who previously indicated last using marijuana within the past 30 days were excluded from the analysis.

<sup>7</sup> This estimate is a composite created from questions MJE02 and MJE03. Past year users who previously indicated having bought the last marijuana the last time they used it were asked question MJE02, and past year users who previously indicated having obtained marijuana another way or not answering the question were asked question MJE03.

<sup>8</sup> Respondents who indicated that they last bought any marijuana within the past 30 days were excluded from the analysis.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.16. Reliability Statistics for Substance Use in the Year Prior to the Past Year among Persons Aged 12 or Older: 2006  
NSDUH Reliability Study**

Substance Use Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Use (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
Cigarettes	3,119	0.16	27.2	95.9 (0.7)	0.90 (0.02)
Alcohol	3,101	0.76	66.4	93.2 (1.0)	0.85 (0.02)
Marijuana and Hashish	3,117	0.29	11.3	95.2 (0.7)	0.75 (0.04)
Cocaine	3,126	0.58	2.7	98.9 (0.2)	0.78 (0.06)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Prior Year Substance Use is defined as using a substance 12 to 23 months prior to the interview date.

NOTE: These questions were asked only of respondents who previously indicated ever using the specified substance. Respondents who reported no lifetime use of a substance were assigned a "no" value to the corresponding prior year use variable even though the question was not asked.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.17. Reliability Statistics for Age at Last Use of Substances among Lifetime Users Aged 12 or Older Who Last Used Respective Substance More Than 30 Days Ago: 2006 NSDUH Reliability Study**

Age at Last Substance Use Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
<b>Cigarettes</b>							
Daily Cigarette Use <sup>5</sup>	864	3.11	52.1 (3.8)	0.51 (0.04)	71.0 (3.2)	0.69 (0.03)	0.13 (0.05)
<b>Snuff</b>	423	2.17	63.9 (4.6)	0.62 (0.05)	81.2 (3.2)	0.80 (0.03)	0.04 (0.02)
<b>Chewing Tobacco</b>	239	0.68	33.4 (7.0)	0.31 (0.07)	75.0 (5.4)	0.72 (0.06)	0.04 (0.02)
<b>Cigars</b>	246	1.13	34.6 (7.4)	0.31 (0.07)	55.6 (8.0)	0.50 (0.09)	0.24 (0.09)
<b>Alcohol</b>	668	3.50	46.8 (4.3)	0.45 (0.04)	64.3 (4.3)	0.62 (0.04)	0.13 (0.04)
<b>Marijuana and Hashish</b>	651	5.59	61.2 (4.7)	0.60 (0.05)	76.0 (4.1)	0.75 (0.04)	0.07 (0.02)
<b>Cocaine</b>	821	1.15	44.8 (3.9)	0.42 (0.04)	67.6 (3.9)	0.63 (0.04)	0.23 (0.13)
<b>Hallucinogens</b>	289	4.36	54.9 (6.0)	0.53 (0.06)	67.7 (6.1)	0.64 (0.07)	0.31 (0.14)
LSD <sup>6</sup>	390	3.44	43.5 (5.3)	0.40 (0.05)	65.4 (6.0)	0.59 (0.06)	0.15 (0.04)
Ecstasy <sup>7</sup>	224	2.94	47.7 (6.2)	0.43 (0.07)	71.7 (5.8)	0.63 (0.08)	0.12 (0.05)
<b>Inhalants</b>	171	1.11	48.7 (8.0)	0.45 (0.08)	71.2 (8.4)	0.63 (0.11)	0.04 (0.01)
<b>Pain Relievers</b>	213	7.33	47.0 (8.1)	0.43 (0.08)	65.3 (7.7)	0.58 (0.09)	0.35 (0.16)
<b>Tranquilizers</b>	276	4.79	38.5 (6.1)	0.37 (0.06)	68.8 (7.2)	0.66 (0.07)	0.13 (0.08)
<b>Stimulants<sup>8</sup></b>	156	2.41	33.5 (7.8)	0.31 (0.07)	40.4 (8.8)	0.37 (0.08)	0.27 (0.10)
Methamphetamine <sup>9</sup>	153	6.94	47.4 (8.0)	0.44 (0.09)	73.0 (7.7)	0.69 (0.09)	0.10 (0.06)
	72	6.28	58.2 (11.7)	0.54 (0.13)	78.7 (11.5)	0.75 (0.13)	0.02 (0.01)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Respondents were instructed to provide their age at last use, in whole years, for each substance used. Questions about age at last use were asked only of respondents who previously indicated last using the specified substance more than 30 days ago. The agreement criterion for approximate agreement was relaxed by 1 year of age.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> Daily Cigarette Use is defined as ever smoking every day for at least 30 days.

<sup>6</sup> This estimate was created based on two questions: One was asked of respondents who reported using LSD only, and the other of respondents who reported using LSD plus another hallucinogen.

<sup>7</sup> This estimate was created based on two questions: One was asked of respondents who reported using Ecstasy only, and the other of respondents who reported using Ecstasy plus another hallucinogen.

<sup>8</sup> This estimate was created as the maximum value of the age at last use of methamphetamine and of other stimulants.

<sup>9</sup> Age at last use of methamphetamine was asked of lifetime users of methamphetamine based on questions only from the core stimulants module.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.18. Reliability Statistics for Number of Months from Last Substance Use to Time of Initial Interview among Recent Substance Initiates Aged 12 or Older: 2006 NSDUH Reliability Study**

Variable for Number of Months from Last Substance Use	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Reporting Consistently (SE)	Kappa (SE)	Weighted Reporting Consistently (SE)	Kappa (SE)	
Cigarettes	146	12.97	67.5 (7.8)	0.64 (0.08)	77.8 (7.0)	0.73 (0.08)	0.36 (0.19)
Cigars	149	15.51	45.1 (8.2)	0.41 (0.08)	62.1 (7.7)	0.55 (0.09)	0.36 (0.09)
Alcohol	345	10.44	60.8 (7.3)	0.54 (0.07)	75.8 (6.7)	0.64 (0.09)	0.28 (0.14)
Marijuana	170	15.65	66.7 (6.1)	0.63 (0.07)	79.9 (5.1)	0.74 (0.07)	0.42 (0.18)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Estimates were created using month and year of last use for each substance, which were asked only of respondents who reported initiating the respective substance at their current age or 1 year younger than their current age. Number of months from last substance use was calculated as the number of months from the month of last use to the month of the initial interview. The agreement criterion for approximate agreement was relaxed by 1 month.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).



**Table 6.19. Reliability Statistics for Substance Use Treatment among Substance Users Aged 12 or Older: 2006 NSDUH Reliability Study**

Substance Use Treatment Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Activity (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
Received Alcohol or Drug Treatment Lifetime <sup>4</sup>	2,290	0.08	7.7	98.5 (0.4)	0.89 (0.03)
Past Year <sup>5</sup>	2,289	0.08	2.0	99.5 (0.1)	0.87 (0.04)
Felt Needed Treatment for Alcohol or Drug Use <sup>6</sup>	2,204	0.03	1.8	99.5 (0.2)	0.86 (0.12)
Enrolled in Treatment Program on October 1st of Last Year <sup>7</sup>	150	1.24	11.5	90.6 (4.6)	0.42 (0.21)
Attended Self-Help Group in the Past Year for Help with Alcohol or Drug Use <sup>8</sup>	2,288	0.21	3.7	98.7 (0.3)	0.81 (0.07)
How Long Since Last Received Treatment for Alcohol/Drugs <sup>9</sup>	119	0.63	N/A	96.4 (1.8)	0.87 (0.07)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

N/A: Not applicable.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> This question was asked only of respondents who previously indicated ever using alcohol or drugs.

<sup>5</sup> This question was asked only of respondents who previously indicated ever using alcohol or drugs and having ever received treatment for alcohol or drug use. Respondents who reported receiving no treatment in lifetime were assigned a "no" value to the corresponding past year variable even though the question was not asked.

<sup>6</sup> This question was asked only of respondents who did not previously indicate ever receiving treatment for alcohol or drugs. This estimate is not comparable with the estimate in NSDUH's 12-month analysis tables for felt need for treatment, which is based on an additional question.

<sup>7</sup> This question was asked only of respondents who previously indicated ever receiving treatment for alcohol or drug use.

<sup>8</sup> This question was asked only of respondents who previously indicated ever using alcohol or drugs and did not previously indicate receiving treatment in the past year in a self-help group. Respondents who indicated receiving treatment in a self-help group in the past year were assigned a "yes" response to this estimate even though the question was not asked. This estimate is not comparable with the estimate in NSDUH's 12-month analysis tables for receipt of treatment in the past year at a self-help group, which is based only on the question asked of respondents who previously indicated receiving treatment in the past year.

<sup>9</sup> Response options for this question are "within the past 30 days," "more than 30 days ago but within the past 12 months," and "more than 12 months ago." This question was asked only of respondents who previously indicated receiving treatment, but did not indicate they were currently receiving treatment. Respondents who reported receiving treatment within the past 30 days or who previously reported that they were currently receiving treatment were excluded from the analysis.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.20. Reliability Statistics for Past Year Health Care and Health Conditions among Persons Aged 12 or Older: 2006  
NSDUH Reliability Study**

Past Year Health Care/Health Condition Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Condition (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Currently Pregnant<sup>4</sup></b>	1,365	1.56	3.4	98.9 (0.7)	0.85 (0.09)
<b>Stayed Overnight as Inpatient in Hospital Treated in the Emergency Room Health Condition<sup>5</sup></b>	3,121	0.23	10.1	95.5 (0.7)	0.76 (0.04)
	3,042	2.47	25.5	87.6 (1.3)	0.67 (0.03)
Anxiety Disorder	3,057	1.42	3.7	98.1 (0.5)	0.73 (0.07)
Asthma	3,055	1.81	6.5	97.3 (0.6)	0.78 (0.05)
Bronchitis	3,059	1.39	3.0	97.9 (0.5)	0.67 (0.07)
Cirrhosis of the Liver	3,060	1.38	0.0	100.0 (0.0)	* (*)
Depression	3,054	1.80	7.7	95.6 (0.8)	0.69 (0.05)
Diabetes	3,058	1.40	7.3	97.3 (0.7)	0.80 (0.06)
Heart Disease	3,060	1.38	6.5	96.1 (1.1)	0.69 (0.10)
Hepatitis	3,060	1.38	0.3	99.8 (0.1)	* (*)
High Blood Pressure	3,058	1.44	21.5	90.1 (1.6)	0.71 (0.04)
HIV/AIDS	3,059	1.46	0.3	100.0 (0.0)	* (*)
Lung Cancer	3,060	1.38	0.7	99.8 (0.2)	* (*)
Pancreatitis	3,060	1.38	0.0	100.0 (0.0)	* (*)
Pneumonia	3,060	1.38	2.7	97.9 (0.7)	0.55 (0.15)
Sexually Transmitted Disease	3,058	1.38	2.1	98.5 (0.5)	0.56 (0.21)
Sinusitis	3,058	1.44	3.8	95.8 (1.3)	0.55 (0.10)
Sleep Apnea	3,060	1.38	2.4	99.5 (0.2)	0.90 (0.04)
Stroke	3,060	1.38	0.2	99.4 (0.4)	* (*)
Tinnitus	3,059	1.39	1.8	99.6 (0.2)	0.88 (0.08)
Tuberculosis	3,060	1.38	0.4	99.7 (0.3)	* (*)
Ulcer or Ulcers	3,059	1.41	1.0	99.5 (0.2)	* (*)
Never Had Any of These Conditions	3,044	2.01	54.1	88.8 (1.5)	0.78 (0.03)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> This question was asked only of females aged 12 to 44.

<sup>5</sup> Respondents were instructed to select all of the health conditions that applied from a list of the conditions they could have had in their lifetime. Respondents who reported "no" to a lifetime condition were assigned a "no" value to the corresponding past year condition even though it was not included as a response option.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.21. Reliability Statistics for Frequency of Past Year Health Care among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Frequency of Health Care Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
<b>Number of Times Been Treated in an Emergency Room<sup>5</sup></b>	3,042	2.47	81.8 (1.4)	0.57 (0.03)	91.2 (1.0)	0.56 (0.05)	0.56 (0.16)
<b>Number of Nights Stayed in a Hospital Overnight or Longer<sup>6</sup></b>							
No Visit Treated as Legitimate Skip <sup>7</sup>	218	1.23	70.2 (7.5)	0.65 (0.08)	87.3 (5.7)	0.81 (0.08)	0.26 (0.20)
No Visit Treated as 0 Visits <sup>8</sup>	3,112	0.33	93.1 (1.0)	0.65 (0.05)	96.1 (0.8)	0.72 (0.05)	0.41 (0.19)
<b>Number of Visits to a Private Therapist<sup>9</sup></b>							
No Visit Treated as Legitimate Skip <sup>10</sup>	83	10.58	42.3 (9.7)	0.38 (0.10)	88.4 (7.5)	0.86 (0.09)	0.04 (0.03)
No Visit Treated as 0 Visits <sup>11</sup>	2,147	0.63	95.4 (0.8)	0.56 (0.05)	98.5 (0.5)	0.80 (0.06)	0.10 (0.05)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> Respondents were instructed to enter the number of times they were treated in an emergency room in the past year. The agreement criterion for approximate agreement was relaxed by one time.

<sup>6</sup> Respondents were instructed to enter the number of nights they stayed in a hospital in the past year. The agreement criterion for approximate agreement was relaxed by 1 night. Respondents who answered that they had not stayed overnight or longer as an inpatient in a hospital to a previous question were excluded from the analysis.

<sup>7</sup> Respondents who answered that they had not stayed overnight or longer as an inpatient in a hospital to a previous question were assigned a value of 0 nights.

<sup>8</sup> Respondents were instructed to enter the number of visits they made to a private therapist in the past year. This question was asked only of respondents who previously indicated having received outpatient treatment at the office of a private therapist, psychologist, psychiatrist, social worker, or counselor who was not part of a clinic. The agreement criterion for approximate agreement was relaxed by five visits.

<sup>9</sup> Respondents who answered that they had not visited a private therapist to a previous question were excluded from the analysis.

<sup>10</sup> Respondents who answered that they had not visited a private therapist to a previous question were assigned a value of 0 visits.

<sup>11</sup> Respondents who answered that they had not visited a private therapist to a previous question were assigned a value of 0 visits.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.22. Reliability Statistics for Lifetime Health Condition Variables among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Lifetime Health Condition Variable <sup>1</sup>	Number of Respondents <sup>2</sup>	Weighted Percent Missing <sup>3</sup>	Percent Reporting Condition (Weighted) <sup>4</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
Anxiety Disorder	3,060	1.38	6.6	97.4 (0.6)	0.78 (0.05)
Asthma	3,060	1.38	12.1	97.1 (0.7)	0.86 (0.03)
Bronchitis	3,060	1.38	9.9	94.9 (1.1)	0.72 (0.05)
Cirrhosis of the Liver	3,060	1.38	0.2	99.9 (0.0)	* (*)
Depression	3,060	1.38	13.0	94.7 (1.0)	0.76 (0.04)
Diabetes	3,060	1.38	8.9	98.4 (0.6)	0.90 (0.04)
Heart Disease	3,060	1.38	9.5	98.9 (0.4)	0.94 (0.03)
Hepatitis	3,060	1.38	0.7	99.6 (0.1)	* (*)
High Blood Pressure	3,060	1.38	27.1	94.1 (1.4)	0.86 (0.03)
HIV/AIDS	3,060	1.38	0.3	100.0 (0.0)	* (*)
Lung Cancer	3,060	1.38	0.7	99.5 (0.4)	* (*)
Pancreatitis	3,060	1.38	0.5	100.0 (0.0)	* (*)
Pneumonia	3,060	1.38	9.1	96.1 (1.0)	0.78 (0.06)
Sexually Transmitted Disease	3,060	1.38	4.6	99.0 (0.3)	0.88 (0.04)
Sinusitis	3,060	1.38	8.6	94.0 (1.5)	0.68 (0.07)
Sleep Apnea	3,060	1.38	3.9	98.9 (0.4)	0.85 (0.06)
Stroke	3,060	1.38	1.0	98.8 (0.5)	0.59 (0.16)
Tinnitus	3,060	1.38	3.5	99.5 (0.2)	0.93 (0.04)
Tuberculosis	3,060	1.38	0.5	99.4 (0.4)	* (*)
Ulcer or Ulcers	3,060	1.38	2.9	97.1 (1.0)	0.53 (0.11)
Never Had Any of These Conditions	3,060	1.38	39.0	94.1 (0.9)	0.87 (0.02)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents were instructed to select all of the health conditions that applied from a list of 20 conditions or the "none of the above" response option.

<sup>2</sup> Respondents with nonmissing data at both interviews.

<sup>3</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>4</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.23. Reliability Statistics for Mental Health Treatment in the Past Year among Persons Aged 18 or Older: 2006 NSDUH Reliability Study**

Mental Health Treatment Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Activity/Location/Source (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Took Prescription Medication for Mental Health Received Outpatient Mental Health Treatment<sup>4</sup></b>	2,157	0.06	12.3	96.8 (0.9)	0.85 (0.04)
<b>Location of Outpatient Treatment<sup>5</sup></b>	2,155	0.08	9.6	97.4 (0.5)	0.85 (0.03)
Outpatient Mental Health Clinic or Center	155	0.12	19.3	84.5 (6.3)	0.43 (0.17)
Office of Private Therapist, Psychologist, Psychiatrist, Social Worker, or Counselor - Not Part of a Clinic	155	0.12	59.5	85.0 (5.4)	0.66 (0.12)
Doctor's Office - Not Part of a Clinic	155	0.12	19.8	83.8 (5.8)	0.52 (0.15)
Outpatient Medical Clinic	155	0.12	5.0	89.7 (5.2)	0.02 (0.08)
Partial Day Hospital or Day Treatment Program	155	0.12	0.9	99.0 (0.9)	0.03 (0.04)
Some Other Place	155	0.12	3.3	99.0 (0.5)	0.86 (0.13)
<b>Source of Payment for Outpatient Treatment<sup>6</sup></b>					
Self or Family Member Living in Household	154	0.14	43.0	85.6 (5.2)	0.71 (0.11)
Family Member Not Living in Household	154	0.14	2.9	99.6 (0.3)	0.93 (0.07)
Private Health Insurance	154	0.14	47.9	91.1 (4.4)	0.82 (0.09)
Medicare	154	0.14	9.8	90.5 (5.2)	0.50 (0.18)
Medicaid	154	0.14	5.4	98.6 (0.9)	0.87 (0.09)
Rehabilitation Program	154	0.14	0.8	99.0 (1.0)	* (*)
Employer	154	0.14	3.5	98.7 (0.6)	0.77 (0.15)
VA or Other Military Program	154	0.14	7.4	100.0 (0.0)	1.00 (0.00)
Other Public Source	154	0.14	0.4	99.7 (0.2)	* (*)
Other Private Source	154	0.14	7.9	90.0 (6.0)	0.00 (0.01)
Free Treatment	154	0.14	1.1	98.5 (0.9)	* (*)
<b>How Prompted to Get Treatment<sup>7</sup></b>	380	4.79	N/A	91.0 (1.5)	0.57 (0.09)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Outpatient mental health treatment/counseling is defined as having received treatment at any of the following locations for problems with emotions, nerves, or mental health: outpatient mental health clinic or center or office of a private therapist, psychologist, psychiatrist, social worker, or counselor that was not part of a clinic.

<sup>5</sup> Respondents were instructed to select all locations of treatment that applied from a list of six options, including "some other place." This question was asked only of respondents who previously indicated having received outpatient treatment in the past year. Estimates may differ from estimates in NSDUH's 12-month analysis tables, which incorporate the other-specify question.

<sup>6</sup> Respondents were instructed to select all sources of payment that applied from a list of 11 options, including "no one paid because the treatment was free." This question was asked only of respondents who previously indicated having received outpatient treatment in the past year.

<sup>7</sup> Response options for this question are "I decided on my own to get treatment," "I got treatment mainly because someone else thought I should," and "I was ordered to get treatment." This question was asked of respondents who had previously indicated having received any kind of mental health treatment in the past year, including inpatient, outpatient, or prescription medication.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.24. Reliability Statistics for Unmet Need for Mental Health Treatment/Counseling in the Past Year among Persons Aged 18 or Older: 2006 NSDUH Reliability Study**

Unmet Need for Mental Health Treatment/Counseling Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Need/Reason (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Needed But Did Not Receive Mental Health Treatment/Counseling</b>	2,154	0.08	5.3	96.5 (0.9)	0.65 (0.08)
<b>Reason for Not Receiving Mental Health Treatment<sup>4</sup></b>					
Could Not Afford Cost	108	0.00	43.8	83.0 (6.4)	0.66 (0.13)
Might Cause Neighbors/Community to Have Negative Opinions	108	0.00	7.7	98.2 (1.2)	0.86 (0.10)
Might Have Negative Effect on Job	108	0.00	9.9	97.7 (1.9)	0.83 (0.13)
Health Insurance Does Not Cover Any Mental Health Treatment/Counseling	108	0.00	4.7	92.1 (3.9)	0.16 (0.10)
Health Insurance Does Not Cover Enough Mental Health Treatment/Counseling	108	0.00	14.4	98.8 (0.7)	0.95 (0.04)
Did Not Know Where to Go for Services	108	0.00	18.0	88.8 (4.7)	0.17 (0.13)
Concerned about Confidentiality	108	0.00	9.4	95.9 (2.2)	0.79 (0.12)
Fear of Being Committed or Having to Take Medicine	108	0.00	23.1	97.0 (2.2)	0.91 (0.07)
Some Other Reason	108	0.00	49.4	73.5 (7.8)	0.47 (0.16)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Mental Health Treatment/Counseling is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Estimates were based only on responses to items in the adult mental health service utilization module.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Respondents were instructed to select all reasons for not receiving treatment that applied from a list of nine options, including "some other reason or reasons." This question was asked only of respondents who previously indicated needing mental health treatment or counseling in the past year, but did not receive it. Estimates may differ from estimates in NSDUH's 12-month analysis tables, which incorporate the other-specify question.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.25. Reliability Statistics for Alternative Mental Health Treatment, Counseling, or Support in the Past Year among Persons Aged 18 or Older: 2006 NSDUH Reliability Study**

Alternative Mental Health Treatment Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Treatment (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Received Alternative Mental Health Treatment</b>	2,154	0.14	9.4	93.8 (1.4)	0.66 (0.07)
<b>Specific Types of Alternative Mental Health Treatment<sup>4</sup></b>					
Acupuncturist/Acupressurist	163	0.17	13.5	97.5 (1.4)	0.90 (0.06)
Chiropractor	163	0.17	37.1	94.7 (3.3)	0.89 (0.07)
Herbalist	163	0.17	4.7	98.0 (1.1)	0.82 (0.11)
In-Person Support Group or Self-Help Group	163	0.17	11.1	95.6 (2.1)	0.80 (0.11)
Internet Support Group or Chat Room	163	0.17	1.4	99.8 (0.1)	0.95 (0.05)
Spiritual or Religious Advisor	163	0.17	23.5	92.6 (2.9)	0.75 (0.10)
Telephone Hotline	163	0.17	1.3	99.9 (0.1)	0.96 (0.05)
Massage Therapist	163	0.17	35.0	93.0 (3.5)	0.85 (0.07)
Other	163	0.17	5.7	96.9 (1.2)	0.60 (0.20)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Respondents were instructed to select all types of alternative mental health treatments received in the past year from a list of nine options, including "other." This question was asked only of respondents who previously indicated receiving alternative types of treatment, counseling, or support for emotions, nerves, or mental health in the past year.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.26. Reliability Statistics for Social Environment Variables among Persons Aged 18 or Older: 2006 NSDUH Reliability Study**

Social Environment Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
<b>Number of Times Performed Behavior in the Past Year<sup>5</sup></b>							
Sold Illegal Drugs	2,152	0.15	98.3 (0.6)	0.43 (0.10)	99.3 (0.3)	0.56 (0.15)	0.28 (0.09)
Stole/Tried to Steal Anything Worth >\$50	2,155	0.07	99.3 (0.2)	0.62 (0.11)	99.8 (0.1)	0.62 (0.20)	0.50 (0.16)
Attacked Someone with Intent to Seriously Hurt	2,156	0.07	99.0 (0.2)	0.55 (0.10)	100.0 (0.0)	0.93 (0.06)	0.18 (0.08)
<b>Feelings about Adults Trying Marijuana Once or Twice<sup>6</sup></b>							
Number of Times Attended Religious Services in the Past Year <sup>7</sup>	2,135	1.02	79.1 (1.9)	0.66 (0.03)	94.3 (1.1)	0.85 (0.03)	0.22 (0.03)
<b>Religious Beliefs<sup>8</sup></b>							
Importance of Beliefs	2,126	0.82	71.4 (2.0)	0.58 (0.03)	91.5 (1.2)	0.69 (0.05)	0.41 (0.05)
Influence of Beliefs on Decisions	2,128	1.00	72.4 (2.4)	0.60 (0.03)	94.2 (1.0)	0.78 (0.04)	0.32 (0.04)
Importance that Friends Share Beliefs	2,128	0.85	65.4 (2.4)	0.49 (0.03)	95.1 (1.5)	0.78 (0.06)	0.33 (0.03)
<b>Number of Times Moved in the Past 5 Years<sup>9</sup></b>	2,142	1.30	84.8 (1.5)	0.77 (0.02)	95.5 (0.9)	0.87 (0.02)	0.33 (0.14)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> Response options for these questions are "0 times," "1 or 2 times," "3 to 5 times," "6 to 9 times," and "10 or more times." The agreement criterion for approximate agreement was relaxed by one category.

<sup>6</sup> Response options for these questions are "neither approve nor disapprove," "somewhat disapprove," and "strongly disapprove." The agreement criterion for approximate agreement was relaxed by one category.

<sup>7</sup> Response options for this question are "0 times," "1 or 2 times," "3 to 5 times," "6 to 24 times," "25 to 52 times," and "more than 52 times." The agreement criterion for approximate agreement was relaxed by one category.

<sup>8</sup> Response options for these questions are "strongly disagree," "disagree," "agree," and "strongly agree." The agreement criterion for approximate agreement was relaxed by one category.

<sup>9</sup> Respondents were instructed to enter the number of times they moved in the past 5 years. The agreement criterion for approximate agreement was relaxed by one move. Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).



**Table 6.27. Reliability Statistics for School Experiences and Religious Beliefs among Persons Aged 12 to 17: 2006 NSDUH Reliability Study**

School Experiences or Religious Beliefs Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	All Response Options (Categorical) <sup>3</sup>		Dichotomized Response <sup>3</sup>		
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Percent Reporting Experience (Weighted) <sup>4</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>School Experiences in the Past Year<sup>5</sup></b>							
Overall Feeling about School <sup>6</sup>	884	0.73	79.8 (2.8)	0.69 (0.04)	80.3	92.9 (1.7)	0.78 (0.06)
How Often Felt Schoolwork Was Meaningful <sup>7</sup>	886	0.02	68.5 (3.0)	0.46 (0.05)	80.4	83.7 (2.3)	0.46 (0.07)
Importance of Things Learned for Later in Life <sup>8</sup>	884	0.10	66.7 (2.9)	0.45 (0.04)	89.0	87.8 (2.2)	0.34 (0.09)
How Interesting Are Courses <sup>9</sup>	885	0.02	70.0 (3.0)	0.53 (0.05)	75.8	85.8 (2.6)	0.64 (0.07)
How Often Teacher Let You Know You Did a Good Job <sup>7</sup>	884	0.04	60.3 (3.1)	0.41 (0.05)	81.8	80.4 (2.7)	0.47 (0.07)
Grades for Last Semester Completed <sup>10</sup>	873	1.56	87.4 (2.0)	0.82 (0.03)	7.0	96.6 (1.3)	0.75 (0.08)
<b>Religious Beliefs<sup>11</sup></b>							
Importance of Beliefs	948	4.97	71.2 (2.8)	0.58 (0.04)	77.5	87.9 (1.6)	0.67 (0.04)
Influence of Beliefs on Decisions	953	3.01	70.9 (2.7)	0.58 (0.04)	72.9	87.8 (2.1)	0.70 (0.05)
Importance That Friends Share Beliefs	941	3.30	68.1 (3.0)	0.56 (0.04)	38.6	86.9 (2.1)	0.72 (0.04)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> The response options for both the All Response Options (Categorical) and the Dichotomized Response columns are specified in footnotes for each of the individual row headers.

<sup>4</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>5</sup> These questions were asked only of youths aged 12 to 17 who attended school or were homeschooled in the past 12 months.

<sup>6</sup> Response options for this question are "you liked going to school a lot," "you kind of liked going to school," "you didn't like going to school very much," and "you hated going to school." The dichotomized response corresponds to persons reporting "you liked going to school a lot" or "you kind of liked going to school."

<sup>7</sup> Response options for this question are "always," "sometimes," "seldom," and "never." The dichotomized response corresponds to persons reporting "always" or "sometimes."

<sup>8</sup> Response options for this question are "very important," "somewhat important," "somewhat unimportant," and "very unimportant." The dichotomized response corresponds to persons reporting "very important" or "somewhat important."

<sup>9</sup> Response options for this question are "very interesting," "somewhat interesting," "somewhat boring," and "very boring." The dichotomized response corresponds to persons reporting "very interesting" or "somewhat interesting."

<sup>10</sup> Response options for this question are "an 'A+', 'A', or 'A-minus' average," "a 'B+', 'B', or 'B-minus' average," "a 'C+', 'C', or 'C-minus' average," "a 'D' or less than a 'D' average," and "my school does not give these grades." The dichotomized response corresponds to persons reporting "a 'D' or less than a 'D' average."

<sup>11</sup> Response options for these questions are "strongly disagree," "disagree," "agree," and "strongly agree." The dichotomized response corresponds to persons reporting "agree" or "strongly agree."

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.28. Reliability Statistics for Perceptions of Cigarette, Alcohol, and Marijuana Use among Persons Aged 12 to 17: 2006  
NSDUH Reliability Study**

Substance Use Perception Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	All Response Options (Categorical) <sup>3</sup>		Dichotomized Response <sup>3</sup>		
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Percent Reporting Experience (Weighted) <sup>4</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>How Parents Would Feel about You<sup>5</sup></b>							
Smoking 1+ Packs Cigarettes/Day	964	1.39	91.1 (1.6)	0.46 (0.07)	93.4	93.1 (1.5)	0.57 (0.08)
Trying Marijuana Once or Twice	965	1.36	88.4 (2.0)	0.45 (0.11)	89.7	89.5 (2.0)	0.49 (0.10)
Using Marijuana Once a Month or More	964	1.41	94.4 (1.2)	0.42 (0.08)	95.2	94.8 (1.2)	0.45 (0.09)
Having 1-2 Alcoholic Drinks Nearly Every Day	963	1.42	86.1 (2.6)	0.33 (0.08)	90.5	88.0 (2.3)	0.41 (0.10)
<b>How You Would Feel about Someone Your Age<sup>6</sup></b>							
Smoking 1+ Packs Cigarettes/Day	963	1.39	77.4 (2.9)	0.40 (0.06)	91.0	88.0 (2.5)	0.34 (0.09)
Trying Marijuana Once or Twice	964	1.35	78.1 (2.8)	0.57 (0.05)	83.3	88.8 (2.1)	0.63 (0.07)
Using Marijuana Once a Month or More	964	1.35	78.4 (3.1)	0.54 (0.06)	85.7	88.3 (2.4)	0.56 (0.08)
Having 1-2 Alcoholic Drinks Nearly Every Day	964	1.39	77.2 (2.8)	0.49 (0.05)	86.7	88.1 (2.3)	0.49 (0.09)
<b>How Close Friends Would Feel about You<sup>6</sup></b>							
Smoking 1+ Packs Cigarettes/Day	956	1.04	77.5 (2.4)	0.51 (0.05)	86.8	89.2 (2.2)	0.55 (0.09)
Trying Marijuana Once or Twice	957	1.61	76.4 (3.2)	0.54 (0.06)	81.3	88.0 (2.7)	0.61 (0.08)
Using Marijuana Once a Month or More	957	1.49	78.9 (3.1)	0.54 (0.06)	83.3	87.4 (2.7)	0.55 (0.09)
Having 1-2 Alcoholic Drinks Nearly Every Day	956	1.50	76.3 (2.7)	0.53 (0.05)	85.6	90.4 (1.6)	0.61 (0.07)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> The response options for both the All Response Options (Categorical) and the Dichotomized Response columns are specified in footnotes for each of the individual row headers.

<sup>4</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>5</sup> Response options for these questions are "neither approve nor disapprove," "somewhat disapprove," and "strongly disapprove." The dichotomized response corresponds to persons reporting "strongly disapprove."

<sup>6</sup> Response options for these questions are "neither approve nor disapprove," "somewhat disapprove," and "strongly disapprove." The dichotomized response corresponds to persons reporting "strongly disapprove" or "somewhat disapprove."

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.29. Reliability Statistics for Delinquent Behavior and Extracurricular Activities in the Past Year among Persons Aged 12 to 17: 2006 NSDUH Reliability Study**

Delinquent Behavior or Extracurricular Activities Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	All Response Options (Categorical) <sup>3</sup>		Dichotomized Response <sup>3</sup>		
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Percent Reporting Experience (Weighted) <sup>4</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Delinquent Behavior</b>							
Argued/Fought with at Least One Parent <sup>5</sup>	960	1.75	62.7 (3.1)	0.52 (0.04)	74.3	90.5 (1.7)	0.74 (0.05)
Got into Serious Fight at School/Work <sup>6</sup>	966	1.04	81.6 (2.5)	0.50 (0.05)	23.9	85.9 (2.0)	0.60 (0.06)
Took Part in Group against Group Fight <sup>6</sup>	968	1.10	82.0 (2.8)	0.41 (0.06)	19.2	86.7 (2.2)	0.54 (0.08)
Carried a Handgun <sup>6</sup>	973	0.10	95.9 (1.2)	0.09 (0.05)	1.8	96.0 (1.2)	0.10 (0.05)
Sold Illegal Drugs <sup>6</sup>	975	0.02	97.0 (1.2)	0.50 (0.18)	2.4	97.2 (1.2)	0.53 (0.18)
Stole/Tried to Steal Anything Worth >\$50 <sup>6</sup>	974	0.09	96.1 (1.1)	0.60 (0.13)	5.3	96.6 (1.2)	0.65 (0.13)
Attacked Someone with Intent to Seriously Hurt <sup>6</sup>	970	0.83	92.9 (1.9)	0.45 (0.09)	6.2	95.4 (1.4)	0.63 (0.12)
<b>Participated in One or Fewer Activities<sup>7</sup></b>	974	0.30	N/A	N/A	11.6	90.4 (1.5)	0.60 (0.08)
School-Based Activities	N/A	N/A	66.6 (3.1)	0.55 (0.04)	N/A	N/A	N/A
Community-Based Activities	N/A	N/A	55.9 (2.7)	0.41 (0.04)	N/A	N/A	N/A
Church or Faith-Based Activities	N/A	N/A	63.3 (3.1)	0.50 (0.04)	N/A	N/A	N/A
Other Activities	N/A	N/A	72.6 (2.7)	0.55 (0.04)	N/A	N/A	N/A
<b>Number of Times Attended Religious Services<sup>8</sup></b>	952	1.35	65.3 (3.0)	0.57 (0.04)	37.3	92.4 (1.6)	0.84 (0.03)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

N/A: Not applicable.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> The response options for both the All Response Options (Categorical) and the Dichotomized Response columns are specified in footnotes for each of the individual row headers.

<sup>4</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>5</sup> Response options for these questions are "0 times," "1 or 2 times," "3 to 5 times," "6 to 9 times," and "10 or more times." The dichotomized response corresponds to persons reporting "0 times," "1 or 2 times," "3 to 5 times," or "6 to 9 times."

<sup>6</sup> Response options for these questions are "0 times," "1 or 2 times," "3 to 5 times," "6 to 9 times," and "10 or more times." The dichotomized response corresponds to persons reporting "1 or 2 times," "3 to 5 times," "6 to 9 times," or "10 or more times."

<sup>7</sup> The number of activities participated in was based on four questions: participation in school-based activities, community-based activities, church or faith-based activities, and other activities. For each question, the response categories include "none," "one," "two," and "3 or more."

<sup>8</sup> Response options for this question are "0 times," "1 or 2 times," "3 to 5 times," "6 to 24 times," "25 to 52 times," and "more than 52 times." The dichotomized response corresponds to persons reporting "25 to 52 times" or "more than 52 times."

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.30. Reliability Statistics for Serious Psychological Distress (SPD) and Components of SPD in the Past Year among Persons Aged 18 or Older: 2006 NSDUH Reliability Study**

SPD Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting SPD (Weighted) <sup>3</sup>	Exact Agreement		Approximate Agreement <sup>4</sup>		Index of Inconsistency (SE) <sup>5</sup>
				Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
Serious Psychological Distress K6 Score <sup>6</sup>	2,128	0.95	10.5	93.4 (1.0)	0.64 (0.05)	N/A	N/A	N/A
During Worst Month in Past Year, Frequency of Feeling <sup>7</sup>	2,128	0.95	N/A	29.6 (2.4)	0.21 (0.02)	79.2 (2.1)	0.63 (0.04)	0.20 (0.03)
Nervous	2,140	0.43	N/A	54.1 (2.8)	0.35 (0.04)	91.4 (1.8)	0.70 (0.06)	0.39 (0.04)
Hopeless	2,145	0.62	N/A	68.8 (2.0)	0.45 (0.03)	93.7 (1.0)	0.75 (0.04)	0.31 (0.04)
Restless or Fidgety	2,142	0.70	N/A	62.1 (2.2)	0.44 (0.03)	92.7 (1.2)	0.73 (0.04)	0.32 (0.03)
So Sad/Depressed that Nothing Could Cheer You Up	2,144	0.63	N/A	67.5 (2.1)	0.44 (0.03)	93.2 (1.1)	0.75 (0.04)	0.31 (0.03)
Everything Was an Effort Down on Yourself, No Good, or Worthless	2,142	0.65	N/A	60.3 (2.8)	0.42 (0.04)	92.2 (1.1)	0.76 (0.03)	0.30 (0.03)
	2,142	0.72	N/A	68.1 (2.1)	0.46 (0.03)	94.0 (1.0)	0.78 (0.03)	0.26 (0.02)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

N/A: Not applicable.

NOTE: Serious Psychological Distress (SPD) is defined as having a score of 13 or higher on the K6 scale. See Section B.4.4 in Appendix B of the *Results from the 2006 National Survey on Drug Use and Health: National Findings*. SPD questions are asked only of adults aged 18 or older.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>5</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>6</sup> The K6 score was calculated by summing across transformed responses ("all of the time" = 4, "most of the time" = 3, "some of the time" = 2, "a little of the time" = 1, and "none of the time" = 0) to the six component questions (see footnote 7). The agreement criterion for approximate agreement was relaxed by three points.

<sup>7</sup> The six variables are the six components of the K6 score. Response options for these questions are "all of the time," "most of the time," "some of the time," "a little of the time," and "none of the time." The agreement criterion for approximate agreement was relaxed by one category.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.31. Reliability Statistics for Major Depressive Episode (MDE) and Methods of Treatment among Persons Aged 18 or Older: 2006 NSDUH Reliability Study**

MDE Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Condition/Action (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Had MDE in Lifetime<sup>4,5</sup></b>	2,137	0.48	14.0	91.8 (1.2)	0.67 (0.05)
<b>MDE Criteria<sup>6</sup></b>					
Depressed Mood	289	1.26	84.2	88.0 (3.6)	0.14 (0.15)
Diminished Interest or Pleasure in Activities	290	1.12	80.5	92.4 (2.4)	0.58 (0.14)
Significant Weight Loss or Gain	265	10.32	88.0	92.5 (2.0)	0.65 (0.11)
Insomnia or Hypersomnia	269	10.71	96.3	92.9 (2.9)	-0.03 (0.02)
Psychomotor Agitation or Retardation	265	10.82	42.4	73.2 (5.6)	0.45 (0.11)
Fatigue or Loss of Energy	270	9.93	95.2	96.6 (1.2)	0.56 (0.25)
Feelings of Worthlessness	269	13.33	55.1	89.8 (2.7)	0.78 (0.06)
Diminished Ability to Think or Concentrate or Indecisiveness	267	10.48	75.0	84.7 (4.5)	0.55 (0.15)
Suicide Ideation	270	9.93	72.5	93.8 (2.2)	0.81 (0.07)
<b>Had MDE in the Past Year<sup>4,7</sup></b>	2,133	0.62	7.1	93.4 (1.3)	0.52 (0.07)
<b>Saw or Talked to Medical Doctor or Other Professional about Depression<sup>8</sup></b>					
In the Past Year	258	0.27	56.8	93.8 (1.8)	0.87 (0.04)
Currently	258	0.27	39.8	90.0 (5.9)	0.79 (0.12)
<b>Took Prescription Medication That Was Prescribed for Depression<sup>8</sup></b>					
In the Past Year	259	0.00	54.3	97.6 (1.0)	0.95 (0.02)
Currently	259	0.00	48.2	98.6 (0.5)	0.97 (0.01)
<b>Specific Professionals Saw or Talked to in the Past Year about Depression<sup>9</sup></b>					
General Practitioner or Family Doctor	106	0.47	56.9	81.0 (10.6)	0.62 (0.21)
Other Medical Doctor	106	0.47	10.5	88.1 (7.3)	0.10 (0.15)
Psychologist	106	0.47	23.0	93.4 (2.8)	0.80 (0.09)
Psychiatrist or Psychotherapist	106	0.47	29.4	91.6 (5.7)	0.81 (0.12)
Social Worker	106	0.47	3.9	92.0 (6.5)	0.46 (0.27)
Counselor	106	0.47	20.0	88.5 (6.1)	0.68 (0.17)
Other Mental Health Professional	106	0.47	3.2	85.9 (10.0)	0.23 (0.24)
Nurse, Occupational Therapist, or Other Health Professional	106	0.47	0.7	99.2 (0.5)	* (*)
Religious or Spiritual Advisor	106	0.47	16.0	93.1 (5.7)	0.79 (0.17)
Herbalist, Chiropractor, Acupuncturist, or Massage Therapist	106	0.47	11.7	87.2 (10.0)	0.28 (0.24)
Another Type of Helping Professional	106	0.47	10.5	87.5 (10.1)	0.22 (0.26)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Major Depressive Episode (MDE) is defined as a period of at least 2 weeks when a person experienced a depressed mood or loss of interest or pleasure in daily activities and had a majority of the symptoms for depression as described in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

<sup>5</sup> This estimate is created based on multiple questions comprising 9 MDE criteria and multiple gatekeeper questions. Respondents who were skipped out of the criteria questions because the gatekeeper questions determined that they did not have MDE were assigned a "no" value even though the questions were not asked.

<sup>6</sup> Each criterion estimate was created based on multiple questions that were asked only of respondents who passed through several gatekeeper questions.

<sup>7</sup> This question was asked only of respondents who had lifetime MDE or met the suicide ideation criterion. Respondents who were classified as not having lifetime MDE were assigned a "no" value to the past year variable even though the question was not asked.

<sup>8</sup> The past year treatment questions were asked only of respondents who had lifetime MDE or met the suicide ideation criterion. The current treatment questions were asked only of respondents who indicated receiving the specific treatment in the past year. Respondents who did not receive past year treatment were assigned a "no" value for current treatment even though the question was not asked.

<sup>9</sup> Respondents were instructed to select all professionals seen or talked to in the past year about depression from a list of 11 options, including "another type of helping professional." This question was asked only of respondents who previously indicated seeing or talking to a medical doctor or other professional in the past year about depression.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.32. Reliability Statistics for Age at First Major Depressive Episode (MDE), Impact of Depression on Person's Life, and Amount Treatment Helped among Persons Aged 18 or Older: 2006 NSDUH Reliability Study**

MDE Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
Age at First MDE <sup>5,6</sup>	245	2.76	46.4 (6.9)	0.45 (0.07)	66.9 (6.5)	0.65 (0.07)	0.24 (0.13)
Severity of Interference in Performing Activities Caused by Depression <sup>7</sup>							
Home Management	132	3.16	62.3 (8.5)	0.46 (0.11)	99.6 (0.3)	0.98 (0.01)	0.24 (0.05)
Ability to Work	130	4.90	58.7 (8.3)	0.47 (0.11)	91.2 (3.5)	0.79 (0.09)	0.22 (0.07)
Ability to Form and Maintain Close Relationships	131	3.28	61.0 (7.8)	0.44 (0.10)	93.2 (3.4)	0.80 (0.10)	0.26 (0.07)
Social Life	131	3.47	61.4 (7.8)	0.45 (0.12)	96.1 (1.9)	0.88 (0.07)	0.21 (0.09)
Number of Days Totally Unable to Work/Do Normal Activities in Past Year <sup>8</sup>	118	7.75	49.3 (9.5)	0.42 (0.08)	58.9 (9.0)	0.47 (0.08)	0.05 (0.02)
Amount Treatment Helped in Past Year <sup>9</sup>							
Prescription Medication	103	1.46	53.4 (11.2)	0.32 (0.17)	97.9 (1.0)	0.91 (0.04)	0.34 (0.11)
Treatment/Counseling	105	9.96	68.5 (9.5)	0.59 (0.11)	97.6 (0.9)	0.94 (0.02)	0.17 (0.04)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> Major Depressive Episode (MDE) is defined as a period of at least 2 weeks when a person experienced a depressed mood or loss of interest or pleasure in daily activities and had a majority of the symptoms for depression as described in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

<sup>6</sup> Respondents were instructed to either enter their exact age or approximate age, in whole years, at the first period of time where depressive symptoms lasted 2 or more weeks for most of the day nearly every day, depending on whether they said they remembered the exact age. These questions were asked only of respondents who were classified as having lifetime MDE or who met the suicide ideation criterion. The agreement criterion for approximate agreement was relaxed by 1 year of age.

<sup>7</sup> These items make up the Sheehan Disability Scale (SDS), which measures the impact of a disorder on a person's life. Respondents were asked to indicate the level of interference for each of the four domains caused by depression on a scale of 0 to 10, where 0 corresponds to "no interference," 1-3 correspond to "mild interference," 4-6 correspond to "moderate interference," 7-9 correspond to "severe interference," and 10 corresponds to "very severe interference." The variables shown are coded with five levels corresponding to the five interference levels. These questions were asked only of respondents who previously indicated having a period of depression for 2 weeks or longer in the past year. The agreement criterion for approximate agreement was relaxed by one category.

<sup>8</sup> Respondents were instructed to enter the number of days they were unable to work or carry out normal activities. This question was asked only of respondents who previously indicated an interference level of 1-10 or responded "don't know" or "refused" for any of the four domains. The agreement criterion for approximate agreement was relaxed by 3 days.

<sup>9</sup> Response options for these questions are "not at all," "a little," "some," "a lot," and "extremely." These questions were asked only of respondents who previously indicated receiving the specified type of treatment for depression in the past year. The agreement criterion for approximate agreement was relaxed by one category.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.33. Reliability Statistics for Youth Mental Health Dichotomous Variables among Persons Aged 12 to 17: 2006 NSDUH Reliability Study**

<b>Youth Mental Health Dichotomous Variable</b>	<b>Number of Respondents<sup>1</sup></b>	<b>Weighted Percent Missing<sup>2</sup></b>	<b>Percent Reporting Event (Weighted)<sup>3</sup></b>	<b>Weighted Percent Reporting Consistently (SE)</b>	<b>Kappa (SE)</b>
<b>Stayed Overnight or Longer in the Past Year Due to Emotional or Behavioral Problem</b>					
Hospital	972	0.44	1.6	97.9 (1.0)	0.36 (0.18)
Residential Treatment Center	972	0.44	0.7	98.3 (1.0)	* (*)
Foster Care or Therapeutic Foster Care Home	972	0.66	0.8	99.0 (0.9)	* (*)
<b>Visited or Received Treatment/ Counseling in the Past Year Due to Emotional or Behavioral Problem</b>					
Partial Day Hospital or Day Treatment Program	968	0.91	1.7	97.5 (1.0)	0.47 (0.25)
Mental Health Clinic or Center	971	0.47	2.6	97.7 (1.1)	0.23 (0.15)
Private Therapist, Psychologist, Psychiatrist, Social Worker, or Counselor	967	0.57	9.2	93.2 (1.8)	0.58 (0.10)
In-Home Therapist, Counselor, or Family Preservation Worker	969	0.52	2.7	95.4 (1.3)	0.06 (0.05)
Pediatrician or Other Family Doctor	968	0.46	3.3	97.5 (1.0)	0.47 (0.16)
Special Education Services While in a Regular Classroom or Placement in a Special Classroom, Special Program, or Special School <sup>4</sup>	875	0.45	5.3	95.0 (1.6)	0.40 (0.17)
School Counselor, School Psychologist, or Regular Meetings with Teachers <sup>4</sup>	871	0.49	11.7	92.4 (1.5)	0.60 (0.09)
<b>Ever Been in Jail or Detention Center</b>	970	0.25	4.8	97.6 (1.0)	0.75 (0.11)
<b>Ever Been in Foster Care<sup>5</sup></b>	964	1.02	3.7	98.9 (0.5)	0.81 (0.11)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> This question was asked only of respondents who previously indicated they were enrolled in any type of school in the past year.

<sup>5</sup> This question was asked only of respondents who did not previously indicate staying overnight or longer in foster care or a therapeutic foster care home in the past year.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.34. Reliability Statistics for Major Depressive Episode (MDE) and Methods of Treatment among Persons Aged 12 to 17: 2006 NSDUH Reliability Study**

Youth MDE Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Condition/ Event (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Had MDE in Lifetime</b> <sup>4,5</sup>	960	1.24	14.5	92.3 (1.6)	0.66 (0.08)
<b>MDE Indicator</b> <sup>6</sup>	88	1.55	82.9	83.5 (10.7)	0.44 (0.32)
<b>MDE Criteria</b> <sup>7</sup>					
Depressed Mood	84	13.37	82.3	96.2 (2.6)	0.82 (0.17)
Diminished Interest or Pleasure in Activities	88	1.55	85.9	96.7 (1.6)	0.85 (0.11)
Significant Weight Loss or Gain	81	21.92	82.0	92.3 (3.2)	0.53 (0.17)
Insomnia or Hypersomnia	81	10.89	86.4	98.6 (0.8)	0.94 (0.05)
Psychomotor Agitation or Retardation	80	22.17	53.9	68.7 (11.8)	0.37 (0.24)
Fatigue or Loss of Energy	81	21.92	93.7	80.5 (10.3)	0.02 (0.06)
Feelings of Worthlessness	82	10.37	65.9	86.0 (7.8)	0.65 (0.20)
Diminished Ability to Think or Concentrate or Indecisiveness	82	10.37	90.4	90.6 (7.5)	0.35 (0.26)
Suicide Ideation	81	10.58	83.0	77.7 (11.5)	0.19 (0.19)
<b>Had MDE in the Past Year</b> <sup>4,8</sup>	954	2.89	8.7	95.7 (0.9)	0.72 (0.07)
<b>Had Depressed Period of 2 Weeks or Longer in the Past Year</b> <sup>9</sup>	70	2.91	66.7	85.5 (5.6)	0.57 (0.19)
<b>Saw or Talked to Medical Doctor or Other Professional about Depression in the Past Year</b> <sup>10,11</sup>	73	1.00	18.9	70.3 (11.9)	0.30 (0.19)
<b>Took Prescription Medication Prescribed for Depression in the Past Year</b> <sup>10</sup>	74	0.83	13.3	100.0 (0.0)	1.00 (0.00)
<b>Remembered Exact Age When First Depression Was Experienced</b> <sup>10</sup>	73	2.71	60.7	59.2 (12.3)	-0.09 (0.16)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Major Depressive Episode (MDE) is defined as a period of at least 2 weeks when a person experienced a depressed mood or loss of interest or pleasure in daily activities and had a majority of the symptoms for depression as described in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*.

<sup>5</sup> This estimate was created based on multiple questions comprising nine MDE criteria and multiple gatekeeper questions. Respondents who were skipped out of the criteria questions because the gatekeeper questions determined that they did not have MDE were assigned a "no" value.

<sup>6</sup> MDE Indicator is a dichotomous variable indicating whether five or more of the nine criteria were met. This variable was computed only for respondents who were asked the criteria questions.

<sup>7</sup> Each criterion estimate was created based on multiple questions, which were asked only of respondents who passed through several gatekeeper questions.

<sup>8</sup> This question was asked only of respondents who had lifetime MDE or met the suicide ideation criterion. Respondents who were classified as not having lifetime MDE were assigned a "no" value to the past year variable even if they responded "yes" to this question or if the question was not asked.

<sup>9</sup> This question was asked only of respondents who had lifetime MDE or met the suicide ideation criterion. It was the basis for past year MDE, but differed because it was asked only among respondents to this question and was not adjusted to lifetime MDE.

<sup>10</sup> This question was asked only of respondents who had lifetime MDE or met the suicide ideation criterion.

<sup>11</sup> This estimate differs from the estimate in NSDUH's 12-month analysis tables because of a high-weight case that gave a "no" response at the first interview.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).



**Table 6.35. Reliability Statistics for Age at First Major Depressive Episode (MDE) and Impact of Depression on a Person's Life among Persons Aged 12 to 17 Who Were Classified as Having Lifetime MDE or Who Met the Suicide Ideation Criterion: 2006 NSDUH Reliability Study**

Youth MDE Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
Age at First MDE <sup>5</sup>	68	4.19	47.7 (12.6)	0.36 (0.15)	80.4 (10.1)	0.63 (0.19)	0.41 (0.12)
How Much Depression Interferes with School Work, Job, or Relationships <sup>6</sup>	74	0.83	60.7 (12.1)	0.49 (0.14)	77.3 (11.7)	0.45 (0.24)	0.41 (0.19)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> Respondents were instructed to enter either their exact age or approximate age, in whole years, at the first period of time when depressive symptoms lasted 2 or more weeks for most of the day nearly every day, depending on whether they said they remembered the exact age. This question was asked only of respondents who were classified as having lifetime MDE or who met the suicide ideation criterion. The agreement criterion for approximate agreement was relaxed by 1 year of age.

<sup>6</sup> Response options for this question are "not at all," "a little," "some," "a lot," and "extremely." This question was asked only of respondents who were classified as having lifetime MDE or who met the suicide ideation criterion. The agreement criterion for approximate agreement was relaxed by one category.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.36. Reliability Statistics for Alcohol Consumption among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Alcohol Consumption Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Event/Condition (Weighted) <sup>3</sup>	Exact Agreement		Approximate Agreement <sup>4</sup>		Index of Inconsistency (SE) <sup>5</sup>
				Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
Number of Drinks Had at Most Recent Alcohol Use <sup>6</sup>	1,295	0.80	N/A	52.6 (3.1)	0.40 (0.04)	87.6 (1.7)	0.65 (0.05)	0.51 (0.13)
Ever Had 5+ Drinks on the Same Occasion <sup>7</sup>	3,123	0.46	50.9	86.6 (2.0)	0.73 (0.04)	N/A	N/A	N/A
Ever Had 4+ Drinks on the Same Occasion <sup>7,8</sup>	1,705	0.34	46.3	87.9 (2.2)	0.76 (0.04)	N/A	N/A	N/A
Age First Had 5+ Drinks on the Same Occasion <sup>9</sup>	1,243	6.47	N/A	45.3 (3.5)	0.39 (0.04)	70.5 (3.1)	0.61 (0.04)	0.25 (0.07)
Age First Had 4+ Drinks on the Same Occasion <sup>8,9</sup>	141	11.49	N/A	60.7 (7.5)	0.56 (0.08)	81.7 (5.2)	0.75 (0.07)	0.29 (0.12)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

N/A: Not applicable.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>5</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>6</sup> Respondents were instructed to enter the number of drinks consumed. This question was asked only of respondents who used alcohol in the past month. The agreement criterion for approximate agreement was relaxed by two drinks.

<sup>7</sup> Respondents who previously reported no lifetime use of alcohol were assigned a "no" value, and respondents who previously had indicated drinking more than five or four drinks, respectively, on the same occasion in the past month were assigned a "yes" value even though the question was not asked.

<sup>8</sup> This question was only asked of females.

<sup>9</sup> Respondents were instructed to provide in whole years their age at first use. These questions were asked of respondents who previously indicated drinking more than five or four drinks, respectively, on the same occasion. The agreement criterion for approximate agreement was relaxed by 1 year of age.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.37. Reliability Statistics for Household Composition and Family Income among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Household Composition and Family Income Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
Number of People Currently Living in Household <sup>5</sup>	3,136	0.00	96.2 (0.9)	0.95 (0.01)	99.0 (0.6)	0.97 (0.02)	0.02 (0.01)
Family Income <sup>6</sup>	2,724	12.06	82.8 (2.4)	0.76 (0.03)	99.8 (0.1)	0.99 (0.00)	0.08 (0.01)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> Respondents were instructed to enter the number of people currently living in the household. The agreement criterion for approximate agreement was relaxed by one person.

<sup>6</sup> Family Income is a four-level variable created based on multiple questions in the income module. Created levels for this variable are "less than \$20,000," "\$20,000-\$49,999," "\$50,000-\$74,999," and "\$75,000 or more." Proxy respondents were allowed to answer questions in the income module in place of the sample member if they could provide more accurate information. No differentiation was made between responses provided by the survey respondent and those provided by the proxy respondent. The agreement criterion for approximate agreement was relaxed by one category.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.38. Reliability Statistics for Health Insurance Coverage among Persons Aged 12 or Older:  
2006 NSDUH Reliability Study**

Health Insurance Coverage Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Coverage (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Current Health Insurance Coverage</b>					
Medicare	3,130	0.08	23.0	96.9 (1.6)	0.91 (0.04)
Medicaid/CHIP <sup>4</sup>	3,117	0.71	9.3	97.0 (0.9)	0.82 (0.05)
Military <sup>5</sup>	3,130	0.61	7.1	98.1 (0.8)	0.86 (0.04)
Private	3,115	0.13	73.1	95.4 (0.9)	0.88 (0.02)
Other <sup>6</sup>	455	1.15	9.2	99.0 (0.6)	0.73 (0.14)
<b>Private Health Insurance<sup>7</sup></b>					
Obtained Insurance through Work	1,965	0.11	88.0	94.2 (1.6)	0.72 (0.07)
Covers Treatment for Alcohol Abuse or Alcoholism	958	50.71	81.2	93.6 (1.9)	0.72 (0.09)
Covers Treatment for Drug Abuse	958	50.32	79.9	91.8 (2.0)	0.66 (0.09)
Covers Treatment for Mental or Emotional Problems	1,282	35.93	91.6	92.6 (2.6)	0.45 (0.12)
<b>Did Not Have Health Insurance/Coverage at Some Time in the Past Year<sup>8</sup></b>	3,106	0.38	17.4	97.2 (0.6)	0.90 (0.02)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Proxy respondents were allowed to answer questions in the health insurance module in place of the sample member if they could provide more accurate information. No differentiation was made between responses provided by the survey respondent and those provided by the proxy respondent.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> CHIP is the Children's Health Insurance Program. Medicaid/CHIP coverage was determined by the combination of two separate questions. The question about Medicaid coverage was asked of all persons, while the question about CHIP coverage was asked only of persons aged 12 to 19.

<sup>5</sup> Military health insurance includes CHAMPUS, TRICARE, CHAMPVA, the VA, and military health care.

<sup>6</sup> This question was asked only of respondents who had previously indicated currently not being covered by Medicare, Medicaid/CHIP, Military, or private health insurance.

<sup>7</sup> These questions were asked only of respondents who previously had indicated currently being covered by private health insurance.

<sup>8</sup> This question was asked only of respondents who previously indicated currently being covered by some type of health insurance. Respondents who reported not currently being covered by health insurance were assigned a "yes" value for this estimate even though the question was not asked.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.39. Reliability Statistics for Lack of Health Insurance Coverage among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Lack of Health Insurance Coverage Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
Number of Months without Health Insurance/Coverage in the Past Year <sup>5</sup>	193	3.57	51.1 (11.9)	0.42 (0.12)	70.4 (11.0)	0.58 (0.14)	0.27 (0.14)
Time Since Last Had Health Care Coverage <sup>6</sup>	424	1.82	67.0 (4.4)	0.57 (0.06)	94.1 (1.6)	0.86 (0.04)	0.18 (0.03)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Proxy respondents were allowed to answer questions in the health insurance module in place of the sample member if they could provide more accurate information. No differentiation was made between responses provided by the survey respondent and those provided by the proxy respondent.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> Respondents were instructed to enter how long, in whole months, they were without health insurance or coverage in the past year. This question was only asked of respondents who previously indicated that they were currently covered by health insurance but were not covered at some point in the past 12 months. The agreement criterion for approximate agreement was relaxed by 1 month.

<sup>6</sup> Response options for this question are "within the past 6 months," "more than 6 months ago, but within the past year," "more than 1 year ago, but within the past 3 years," "more than 3 years ago," and "never had coverage." This question was asked only of respondents who previously indicated not currently being covered by any health insurance. The agreement criterion for approximate agreement was relaxed by one category.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.40. Reliability Statistics for Main Reason Stopped Being Covered by Health Insurance among Persons Aged 12 or Older Who Were Currently without Health Care Coverage But Had Health Care Coverage in the Past: 2006 NSDUH Reliability Study**

Main Reason for Lack of Health Insurance Coverage Variable <sup>1</sup>	Number of Respondents <sup>2</sup>	Weighted Percent Missing <sup>3</sup>	Percent Reporting Reason (Weighted) <sup>4</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
Insured Person Lost Job or Changed Employers	383	1.41	18.0	85.8 (3.2)	0.53 (0.09)
Lost Medicaid/Medical Assistance Coverage Due to New Job or Increased Income	383	1.41	4.8	94.0 (1.6)	0.46 (0.13)
Lost Medicaid/Medical Assistance Coverage for Some Other Reason	383	1.41	9.1	94.6 (1.7)	0.70 (0.10)
Cost Too High/Can't Afford Premiums	383	1.41	27.3	84.8 (4.1)	0.61 (0.10)
Became Ineligible Due to Age or Leaving School	383	1.41	13.8	94.2 (2.0)	0.76 (0.08)
Employer Does Not Offer Coverage/Not Eligible for Coverage	383	1.41	13.7	89.1 (3.9)	0.42 (0.15)
Got Divorced or Separated from Insured Person	383	1.41	1.5	99.7 (0.2)	0.92 (0.07)
Death of Spouse or Parent	383	1.41	0.1	100.0 (0.0)	* (*)
Insurance Company Refused Coverage	383	1.41	2.5	97.9 (1.8)	0.17 (0.18)
Don't Need It	383	1.41	2.3	95.9 (2.0)	0.47 (0.21)
Received Medicaid/Medical Insurance Only While Pregnant	383	1.41	4.0	99.4 (0.4)	0.93 (0.05)
Some Other Reason	383	1.41	3.0	96.1 (1.8)	0.21 (0.14)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Proxy respondents were allowed to answer questions in the health insurance module in place of the sample member if they could provide more accurate information. No differentiation was made between responses provided by the survey respondent and those provided by the proxy respondent.

<sup>1</sup> Respondents were instructed to select one main reason for stopping health coverage among 12 response options. Dichotomous variables were created for each of these 12 response options. This question was asked only among respondents who previously indicated not currently being covered by health insurance, but did not report that they never had coverage.

<sup>2</sup> Respondents with nonmissing data at both interviews.

<sup>3</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>4</sup> Reported rates were provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.41. Reliability Statistics for Types of Income Received among Persons Aged 12 or Older, by Income Sample: 2006 NSDUH Reliability Study**

Income Source Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Income Type (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>SAMPLE A (n = 1,580)</b>					
<b>Social Security or Railroad Retirement Payments</b>					
Sample Member	1,572	0.43	29.4	96.8 (1.2)	0.92 (0.03)
Family Members	1,127	6.54	14.9	97.9 (0.7)	0.92 (0.03)
<b>Supplemental Security Income</b>					
Sample Member	1,564	0.58	3.8	97.0 (1.2)	0.64 (0.12)
Family Members	1,181	6.81	2.2	98.4 (0.5)	0.68 (0.14)
<b>Income from Wages or Pay Working Job/Business</b>					
Sample Member	1,574	0.47	62.0	97.4 (0.6)	0.95 (0.01)
Family Members	535	4.81	59.9	88.3 (6.0)	0.74 (0.13)
<b>Food Stamps</b>					
Sample Member	289	1.38	7.1	99.7 (0.2)	0.98 (0.02)
Family Members	1,263	7.51	11.3	97.9 (0.6)	0.88 (0.03)
<b>Cash Assistance from State/County Welfare</b>					
Sample Member	1,572	0.41	0.8	99.0 (0.3)	* (*)
Family Members	1,189	7.95	0.2	99.2 (0.5)	* (*)
<b>Other Noncash Public Assistance or Welfare<sup>4</sup></b>					
Sample Member	1,568	0.55	2.0	98.1 (0.5)	0.59 (0.09)
Family Members	1,164	8.16	0.8	99.0 (0.5)	* (*)
<b>Interest from Bank Account, Dividend Income from Stocks/Mutual Funds, or Income from Rental Properties, Royalties, Estates, or Trusts</b>					
Sample Member	1,564	1.31	66.1	88.7 (1.6)	0.76 (0.03)
Family Members	510	21.05	28.3	78.7 (5.6)	0.45 (0.12)
<b>Child Support Payments</b>					
Sample Member	1,576	0.40	3.6	99.6 (0.2)	0.94 (0.03)
Family Members	1,192	8.18	4.4	99.2 (0.2)	0.91 (0.03)
<b>Other Income Source</b>					
Sample Member	1,569	0.51	16.2	90.6 (2.3)	0.65 (0.09)
Family Members	1,174	6.96	13.9	92.6 (1.9)	0.71 (0.08)
<b>SAMPLE B (n = 1,556)</b>					
<b>Supplemental Security Income</b>					
	1,532	0.32	7.5	96.6 (0.8)	0.74 (0.07)
<b>Income from Wages or Pay Working Job/Business</b>					
	1,550	0.22	77.7	95.5 (1.0)	0.87 (0.03)
<b>Food Stamps</b>					
	1,550	0.06	8.0	97.8 (0.6)	0.85 (0.04)
<b>Cash Assistance from State/County Welfare</b>					
	1,548	0.34	2.9	98.4 (0.6)	0.69 (0.09)
<b>Noncash Public Assistance or Welfare<sup>4</sup></b>					
	1,553	0.85	3.0	97.6 (0.7)	0.57 (0.10)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: In 2006, approximately half of the reliability respondents (sample A) received the standard income questions, and the other half (sample B) received a new set of income questions. The new questions included a reduced number of items and only family-level questions.

NOTE: Proxy respondents were allowed to answer questions in the income module in place of the sample member if they could provide more accurate information. No differentiation was made between responses provided by the survey respondent and those provided by the proxy respondent.

NOTE: The "sample member" refers to the person selected into the survey who was the respondent for most or all of the questionnaire, and "family members" refer to members of that person's family. In the case of self-reporting, the respondent for the income module answered questions about himself for sample member and others in the family for family members. In the case of proxy-reporting, the respondent for the income module answered questions about another person for the sample member and himself and possibly others for family members.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Noncash public assistance or welfare includes help getting a job, placement in education or job training programs, or help with transportation, child care, or housing.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (n=3,136).

**Table 6.42. Reliability Statistics for Demographics (Dichotomized) among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Demographic (Dichotomized) Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Characteristic (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>General Demographics</b>					
Gender <sup>4</sup>	3,136	0.00	48.1	99.9 (0.1)	1.00 (0.00)
Hispanic, Latino, or Spanish Origin or Descent	3,125	0.10	7.1	99.9 (0.0)	0.99 (0.00)
Ever Been in U.S. Armed Forces	2,288	0.00	17.3	99.9 (0.1)	1.00 (0.00)
Leave Home/Stay in Temporary Housing Due to 2005 Hurricanes <sup>5</sup>	30	0.00	20.9	100.0 (0.0)	1.00 (0.00)
Born in the United States	3,136	0.00	94.5	99.6 (0.2)	0.96 (0.02)
Lived in the United States at Least 1 Year	177	0.00	95.2	96.7 (3.0)	0.31 (0.23)
<b>Education</b>					
Currently Enrolled in Any School	3,136	0.00	16.2	98.6 (0.7)	0.95 (0.02)
Currently on Holiday/Vacation from School <sup>6</sup>	490	0.00	5.4	96.2 (1.5)	0.59 (0.14)
Enrolled in School Full-Time <sup>7</sup>	1,382	0.00	80.6	97.0 (0.9)	0.89 (0.04)
Received High School Diploma <sup>8</sup>	328	0.00	65.2	96.9 (1.2)	0.93 (0.03)
Received GED <sup>9</sup>	104	0.00	19.5	92.6 (4.7)	0.76 (0.14)
<b>Employment<sup>10</sup></b>					
Written Policy at Work about Employee Use of Drugs or Alcohol	1,498	9.29	78.8	92.2 (1.6)	0.78 (0.04)
Educational Information about Drugs or Alcohol Given at Work	1,620	1.51	40.9	82.5 (2.0)	0.65 (0.04)
Access to Employee Assistance Program for Employees with Drug or Alcohol Problem	1,283	18.97	50.2	88.1 (3.0)	0.76 (0.06)

\* Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).  
N/A: Not applicable.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> The estimate shown in the Percent Reporting Characteristic column is for the response category "male."

<sup>5</sup> This question was asked only of residents/former residents of Alabama, Florida, Louisiana, and Mississippi concerning Hurricane Katrina and residents/former residents of Louisiana and Texas concerning Hurricane Rita.

<sup>6</sup> This question was asked only of respondents aged 12 to 25 who had completed 3 years of college or less of school and were not currently enrolled in school or had missing information about current school enrollment.

<sup>7</sup> Response options for this question are "full-time" and "part-time." This question was asked only of respondents who were currently enrolled in school or who were planning to return to school after holiday/vacation.

<sup>8</sup> This question was asked only of respondents aged 12 to 25 who were not currently enrolled in school and were not planning on returning to school or had missing enrollment information, and who had completed 12th grade or less of school or had missing information for highest grade completed.

<sup>9</sup> This question was asked only of respondents who did not receive a high school diploma or had missing information as to whether they received a high school diploma.

<sup>10</sup> These questions were asked only of respondents aged 15 or older who were employed.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).



**Table 6.43. Reliability Statistics for Demographic Scaled Variables among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Demographic Scaled Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Exact Agreement		Approximate Agreement <sup>3</sup>		Index of Inconsistency (SE) <sup>4</sup>
			Weighted Percent Reporting Consistently (SE)	Kappa (SE)	Weighted Percent Reporting Consistently (SE)	Kappa (SE)	
<b>General Demographics</b>							
Number of Times Married <sup>5</sup>	2,604	0.15	98.0 (0.8)	0.97 (0.01)	N/A	N/A	0.02 (0.01)
Highest Grade Completed <sup>6</sup>	3,136	0.00	91.8 (1.2)	0.90 (0.01)	97.6 (0.7)	0.97 (0.01)	0.07 (0.03)
Overall Health <sup>7</sup>	3,135	0.00	76.0 (2.0)	0.67 (0.03)	97.4 (0.6)	0.91 (0.02)	0.17 (0.02)
Number of Times Moved in the Past Year <sup>8</sup>	3,134	0.03	88.6 (1.3)	0.71 (0.03)	97.5 (0.5)	0.80 (0.04)	0.29 (0.05)
Number of Years Lived in the United States <sup>9</sup>	177	0.00	75.1 (5.9)	0.74 (0.06)	94.4 (3.1)	0.94 (0.03)	0.00 (0.00)
<b>Education</b>							
Current Grade Level <sup>6,10</sup>	1,380	0.08	94.7 (1.2)	0.94 (0.01)	99.9 (0.1)	1.00 (0.00)	0.00 (0.00)
Number Days of School Missed Because Sick/Injured <sup>11</sup>	927	0.03	72.5 (3.3)	0.47 (0.04)	96.6 (1.2)	0.68 (0.10)	0.38 (0.09)
Number Days of School Skipped Because Didn't Want to Go <sup>11</sup>	928	0.17	83.6 (3.0)	0.47 (0.07)	98.2 (1.1)	0.71 (0.15)	0.35 (0.08)
Age When Stopped Attending School <sup>12</sup>	103	0.33	76.9 (7.3)	0.69 (0.10)	96.4 (3.0)	0.89 (0.09)	0.15 (0.06)
<b>Employment</b>							
Number of People Who Work for Employer <sup>13</sup>	1,635	0.10	82.0 (2.5)	0.77 (0.03)	93.0 (2.2)	0.85 (0.05)	0.13 (0.03)
What Written Policy Covers <sup>14</sup>	1,011	1.11	94.4 (1.3)	0.31 (0.14)	N/A	N/A	0.71 (0.12)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).  
N/A: Not applicable.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified elsewhere in the footnotes to this table.

<sup>4</sup> The index of inconsistency, a measure of the ratio of the response variance to the total variance, reflects the degree of inconsistency in the responses. See Appendix B for more information.

<sup>5</sup> Respondents were instructed to enter the number of times they have been married. Respondents who answered a previous question that they had never been married were assigned a value of 0 times. Approximate agreement was not used for this variable.

<sup>6</sup> Response options for this question are "1st grade," "2nd grade," "3rd grade," ..., "college or university/3rd year," "college or university/4th year," and "college or university/5th or higher year." The agreement criterion for approximate agreement was relaxed by one grade level.

<sup>7</sup> Response options for this question are "excellent," "very good," "good," "fair," and "poor." The agreement criterion for approximate agreement was relaxed by one category.

<sup>8</sup> Respondents were instructed to enter the number of times moved in the past year. The agreement criterion for approximate agreement was relaxed by one move.

<sup>9</sup> Respondents were instructed to enter in whole years how long they had lived in the United States. Respondents who answered that they had not lived in the United States for at least 1 year to a previous question were assigned a value of 0 years. The agreement criterion for approximate agreement was relaxed by 1 year.

<sup>10</sup> This question was asked only of respondents who were currently enrolled in school or who were planning to return to school after holiday/vacation.

<sup>11</sup> This question was asked only of respondents who were currently or would be a full-time student. The agreement criterion for approximate agreement was relaxed by 3 days.

<sup>12</sup> Respondents were instructed to enter their age, in whole years, when they stopped attending school. This question was asked only of respondents aged 12 to 25 who were not currently enrolled in school and were not planning on returning to school or had missing enrollment information, and did not receive a high school diploma or had missing information as to whether a diploma was received. The agreement criterion for approximate agreement was relaxed by 1 year of age.

<sup>13</sup> Response options for this question are "less than 10 people," "10-24 people," "25-99 people," "100-499 people," and "500 people or more." The agreement criterion for approximate agreement was relaxed by one category.

<sup>14</sup> Response options for this question are "only alcohol," "only drugs," and "both alcohol and drugs." This question was asked only among respondents who said that there was a written policy about employee use of alcohol or drugs at their workplace. Approximate agreement was not used for this variable.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.44. Reliability Statistics for Ethnicity, Race, and Marital Status (Dichotomized) among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Ethnicity, Race, and Marital Status (Dichotomized) Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Characteristic (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>Hispanic, Latino, or Spanish Origin or Descent<sup>4</sup></b>	3,125	0.10	7.1	99.9 (0.0)	0.99 (0.00)
Mexican <sup>5</sup>	3,125	0.10	3.4	99.9 (0.0)	0.98 (0.01)
Puerto Rican <sup>5</sup>	3,125	0.10	1.4	99.9 (0.0)	0.98 (0.01)
Central or South American <sup>5</sup>	3,125	0.10	0.3	99.9 (0.1)	* (*)
Cuban/Cuban American <sup>5</sup>	3,125	0.10	0.5	100.0 (0.0)	* (*)
Dominican <sup>5</sup>	3,125	0.10	0.1	100.0 (0.0)	* (*)
Spanish <sup>5</sup>	3,125	0.10	1.3	99.9 (0.0)	0.97 (0.03)
<b>Race<sup>6</sup></b>					
White	3,132	0.20	84.3	98.8 (0.4)	0.95 (0.01)
Black/African American	3,132	0.20	10.3	99.9 (0.1)	0.99 (0.00)
American Indian or Alaska Native	3,132	0.20	1.1	98.9 (0.4)	0.54 (0.12)
Native Hawaiian	3,132	0.20	0.1	100.0 (0.0)	* (*)
Other Pacific Islander	3,132	0.20	0.3	99.6 (0.2)	* (*)
Asian	3,132	0.20	3.3	99.8 (0.1)	0.97 (0.02)
<b>Currently Married<sup>7</sup></b>	2,605	0.12	49.1	98.7 (0.4)	0.97 (0.01)

\* Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Other-specify responses from question QD05OTHR ("which other racial group describes you?") were incorporated into this estimate.

<sup>5</sup> Respondents were instructed to choose all that apply from the six ethnicity groups and an option for "other." This question was only asked of residents who had previously answered that they are of Hispanic, Latino, or Spanish origin or descent. Other-specify responses were not incorporated into this estimate. Each of the six ethnicity responses was treated as a separate dichotomous variable.

<sup>6</sup> Respondents were instructed to choose all that apply from the six race groups and an option for "other." Other-specify responses were not incorporated into this estimate. Race estimates shown in this table are not mutually exclusive. Estimates here differ from estimates in NSDUH's 12-month analysis tables, which are mutually exclusive and only among non-Hispanics. Each of the six race responses was treated as a separate dichotomous variable.

<sup>7</sup> Response options for this question are "married," "widowed," "divorced or separated," and "have never married." This estimate corresponds to persons reporting "married." This question was asked only of respondents aged 15 or older.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.45. Reliability Statistics for Noncore Demographic Categorical Variables among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Noncore Demographic Categorical Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Percent Reporting Characteristic (Weighted) <sup>3</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>General Demographics</b>					
State of Residence 1 Year Ago Today <sup>4</sup>	854	0.01	N/A	99.0 (0.3)	0.99 (0.00)
State of Residence 2 Years Ago Today <sup>5</sup>	1,715	0.00	N/A	99.0 (0.3)	0.99 (0.00)
Country or U.S. Territory of Birth <sup>6</sup>	176	0.22	N/A	91.7 (4.6)	0.91 (0.05)
<b>Education</b>					
Reason for Leaving School before Receiving High School Diploma <sup>7</sup>	103	3.04	N/A	80.4 (6.4)	0.77 (0.07)
<b>Employment<sup>8</sup></b>					
Current Business/Industry <sup>9</sup>	1,635	0.87	N/A	85.2 (2.2)	0.77 (0.03)
Workplace Tests for Alcohol Usage	1,490	8.31	30.6	88.2 (2.0)	0.72 (0.05)
Workplace Tests for Drug Usage	1,529	7.03	45.9	92.7 (1.3)	0.85 (0.03)
Workplace Tests during Hiring Process <sup>10</sup>	634	2.30	82.5	94.3 (3.2)	0.80 (0.11)
Workplace Tests on Random Basis <sup>10</sup>	594	4.70	54.9	85.8 (3.8)	0.70 (0.08)
What Happens on First Positive Test <sup>10,11</sup>	533	14.25	N/A	76.8 (5.1)	0.65 (0.07)
Would You Work for Employer Who Tests during Hiring Process <sup>12</sup>	1,638	0.22	N/A	82.8 (2.4)	0.68 (0.04)
Would You Work for Employer Who Tests Randomly <sup>12</sup>	1,640	0.31	N/A	79.8 (2.6)	0.65 (0.04)

\* Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

N/A: Not applicable.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

<sup>4</sup> Response options for this question are "each of the 50 States," "the District of Columbia," and "outside of U.S." This question was asked only of respondents who previously indicated having moved at least once in the past year.

<sup>5</sup> Response options for this question are "each of the 50 States," "the District of Columbia," and "outside of U.S." This question was asked only of respondents who previously indicated having moved at least once in the past 5 years.

<sup>6</sup> Respondents were instructed to enter the country or U.S. territory in which they were born. The question was asked only of respondents who previously indicated that they were not born in the United States.

<sup>7</sup> There were 15 response options for this question, including "school was boring or I didn't want to be there," "I got pregnant/I got someone pregnant," and "I got in trouble or expelled for selling drugs." This question was asked only of respondents aged 12 to 25 not enrolled in school who previously indicated that they had not received a high school diploma or who had missing information as to whether they had received a diploma.

<sup>8</sup> These questions were asked only of respondents aged 15 or older who were currently employed.

<sup>9</sup> The response options for this question are "manufacturing," "wholesale trade," "retail trade," "agriculture," "construction," "service," "government," and "other."

<sup>10</sup> This question was asked only of respondents who previously indicated that their workplace tests employees for alcohol or drug use.

<sup>11</sup> Response options for this question are "handled on an individual basis/policy does not specify what happens," "employee is fired," "employee is referred for treatment or counseling," "nothing happens," and "something else happens."

<sup>12</sup> Response options for this question are "more likely," "less likely," and "would make no difference."

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.46. Reliability Statistics for Industry and Occupation Codes among Persons Aged 15 or Older Who Were Ever Employed: 2006 NSDUH Reliability Study**

Industry and Occupation Variable	Number of Respondents <sup>1</sup>	Weighted Percent Missing <sup>2</sup>	Weighted Percent Reporting Consistently (SE)	Kappa (SE)
<b>4-Digit Code<sup>3</sup></b>				
Industry	1,946	0.02	70.3 (2.5)	0.70 (0.03)
Occupation	1,946	0.02	74.1 (2.2)	0.74 (0.02)
<b>2-Digit Code<sup>4</sup></b>				
Industry <sup>5</sup>	1,946	0.02	83.9 (2.0)	0.83 (0.02)
Occupation <sup>6</sup>	1,946	0.02	84.9 (1.7)	0.84 (0.02)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: The reliability measures of the industry and occupation codes were assessed by combining respondents' employment information from both current job and last job. The "not reported/not codable" category was treated as a valid response.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> The 4-digit codes were determined using the Census Bureau's standard industry and occupation coding system.

<sup>4</sup> The 2-digit codes are the first 2 digits of the 4-digit codes.

<sup>5</sup> The categories for the major industry code include "agriculture, forestry, fishing and hunting"; "mining"; "utilities"; "construction"; "manufacturing"; "wholesale trade"; "retail trade"; "transportation and warehousing"; "information including publishing industries and broadcasting/telecommunications"; "finance and insurance"; "real estate and rental and leasing"; "professional, scientific, and technical services"; "management of companies and enterprises, administrative, support, waste management, and remediation services"; "educational services"; "health care and social assistance"; "arts, entertainment, and recreation"; "accommodations and food services"; "other services (except public administration)"; "public administration"; and "armed forces."

<sup>6</sup> The categories for the major occupation code include "management occupations"; "financial occupations"; "mathematical and computer scientists"; "engineering and related technicians, architects, and surveyors"; "life, physical, and social science occupations"; "community and social services occupations: include counselors, social, and religious workers"; "legal occupations"; "education, training and library occupations"; "arts, design, entertainment, sports and media occupations"; "health care practitioners and technical occupations"; "protective service occupations"; "food preparation and serving related occupations"; "building and grounds cleaning and maintenance occupations"; "entertainment attendants and related workers"; "personal care and service occupations"; "sales and related occupations"; "office and administrative support occupations"; farming, fishing, and forestry occupations"; "construction and extraction occupations"; "installation, maintenance, and repair occupations"; "production and operating services, including production and operating workers, as well as setters, operators, and tenders"; "transportation and material moving occupations"; and "armed forces."

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.47. Reliability Statistics for Nonmedical Use of Specific Pain Relievers in the Lifetime among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

<b>Pain Reliever Variable</b>	<b>Number of Respondents<sup>1</sup></b>	<b>Weighted Percent Missing<sup>2</sup></b>	<b>Percent Reporting Use (Weighted)<sup>3</sup></b>	<b>Weighted Percent Reporting Consistently (SE)</b>	<b>Kappa (SE)</b>
Codeine	3,121	0.13	2.9	98.5 (0.4)	0.73 (0.06)
Demerol <sup>®</sup>	3,121	0.13	2.0	99.2 (0.4)	0.78 (0.13)
Dilaudid <sup>®</sup>	3,121	0.13	0.3	100.0 (0.0)	* (*)
Fioricet <sup>®</sup>	3,121	0.13	0.1	99.9 (0.1)	* (*)
Fiorinal <sup>®</sup>	3,121	0.13	1.1	99.1 (0.8)	* (*)
Hydrocodone	3,121	0.13	3.1	98.4 (0.4)	0.73 (0.06)
Methadone	3,121	0.13	0.5	99.6 (0.1)	* (*)
Morphine	3,121	0.13	1.1	99.3 (0.2)	0.70 (0.10)
OxyContin <sup>®</sup>	3,121	0.13	1.4	98.9 (0.4)	0.67 (0.09)
Phenaphen <sup>®</sup> with Codeine	3,121	0.13	0.8	99.3 (0.4)	* (*)
Propoxyphene	3,121	0.13	0.5	99.6 (0.3)	* (*)
SK-65 <sup>®</sup>	3,121	0.13	0.1	100.0 (0.0)	* (*)
Stadol <sup>®</sup>	3,121	0.13	0.0	100.0 (0.0)	* (*)
Talacen <sup>®</sup>	3,121	0.13	0.0	100.0 (0.0)	* (*)
Talwin <sup>®</sup>	3,121	0.13	0.2	99.9 (0.1)	* (*)
Talwin NX <sup>®</sup>	3,121	0.13	0.0	100.0 (0.0)	* (*)
Tramadol	3,121	0.13	0.2	100.0 (0.0)	* (*)
Ultram <sup>®</sup>	3,121	0.13	0.2	99.7 (0.1)	* (*)

\*Suppression of estimated kappa and its associated standard error (SE) according to the suppression rule (see Appendix B).

NOTE: Respondents were instructed to select any of the pain relievers that had ever been used that were not prescribed from a list of 18 specific pain relievers. Respondents who reported no lifetime use of pain relievers were assigned "no" values for lifetime use of each of the specific pain relievers even though this question was not asked.

<sup>1</sup> Respondents with nonmissing data at both interviews.

<sup>2</sup> Does not count legitimate skips where a response level could not be assigned based on responses to previous questions.

<sup>3</sup> Reported rates are provided because of the kappa statistic's dependence on prevalence. Kappa statistics should not be compared when their associated prevalence rates are dissimilar.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.48. Response Propensity Model for Completion of Initial Interview by Sample Members Aged 12 or Older: 2006 NSDUH Reliability Study**

Completion of Initial Interview Variable	Odds Ratio	Coefficient (SE)	p Value	Wald F	p Value (Wald F)
<b>Intercept</b>	2.05	0.72 (0.39)	0.07		
<b>Age</b>				33.37	0.00
12 to 17	2.51	0.92 (0.12)	0.00		
18 to 25	1.63	0.49 (0.11)	0.00		
26 or Older	1.00	0.00 (0.00)	.		
<b>Gender</b>				0.09	0.76
Male	1.03	0.03 (0.11)	0.76		
Female	1.00	0.00 (0.00)	.		
<b>Race/Ethnicity</b>				4.43	0.00
White	1.00	0.00 (0.00)	.		
Black or African American	1.17	0.15 (0.20)	0.44		
Other	0.63	-0.46 (0.25)	0.06		
Hispanic or Latino	1.77	0.57 (0.21)	0.01		
<b>Tract-Level Characteristics</b>					
<b>Region</b>				0.66	0.58
Northeast	1.18	0.16 (0.20)	0.42		
Midwest	1.27	0.24 (0.17)	0.18		
South	1.21	0.19 (0.17)	0.26		
West	1.00	0.00 (0.00)	.		
<b>Population Density</b>				1.28	0.28
Segment in a CBSA with ≥ 1 Million Persons	0.76	-0.27 (0.36)	0.44		
Segment in a CBSA with 250,000 to 999,999 Persons	1.19	0.17 (0.33)	0.60		
Segment in a CBSA with < 250,000 Persons	1.18	0.17 (0.32)	0.60		
Segment Not in a CBSA and Not in a Rural Area	1.15	0.14 (0.45)	0.76		
Segment Not in a CBSA and in a Rural Area	1.00	0.00 (0.00)	.		
<b>Urbanicity</b>				0.92	0.43
Rural	1.00	0.00 (0.00)	.		
Non-MSA Urban	0.92	-0.08 (0.28)	0.76		
MSA < 1 Million	0.70	-0.35 (0.26)	0.17		
MSA ≥ 1 Million	0.81	-0.21 (0.29)	0.47		
<b>Percent Families below Poverty Level</b>	1.95	0.67 (0.89)	0.45	0.56	0.45
<b>Percent Housing Units Rented</b>	1.08	0.07 (0.41)	0.86	0.03	0.86
<b>Percent Bachelor, Graduate, or Professional Degrees</b>	1.36	0.31 (0.62)	0.62	0.25	0.62
<b>Median Household Income (in Multiples of \$10,000)</b>	1.04	0.04 (0.06)	0.48	0.50	0.48

CBSA = core-based statistical area; MSA = metropolitan statistical area; SE = standard error.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (n = 3,136).

**Table 6.49. Response Propensity Model for Completion of Reinterviews by Sample Members Aged 12 or Older: 2006 NSDUH Reliability Study**

Completion of Reinterview Variable	Odds Ratio	Coefficient (SE)	p Value	Wald F	p Value (Wald F)
<b>Intercept</b>	16.82	2.82 (0.57)	0.00		
<b>Age</b>				17.69	0.00
12 to 17	3.37	1.21 (0.21)	0.00		
18 to 25	1.60	0.47 (0.18)	0.01		
26 or Older	1.00	0.00 (0.00)	.		
<b>Gender</b>				0.88	0.35
Male	0.85	-0.17 (0.18)	0.35		
Female	1.00	0.00 (0.00)	.		
<b>Race/Ethnicity</b>				1.65	0.18
White	1.00	0.00 (0.00)	.		
Black or African American	0.83	-0.19 (0.34)	0.57		
Other	1.18	0.17 (0.44)	0.70		
Hispanic or Latino	0.52	-0.66 (0.32)	0.04		
<b>Reported Mental Health Problems at T1</b>				1.56	0.21
Yes	1.39	0.33 (0.26)	0.21		
No	1.00	0.00 (0.00)	.		
<b>Reported Lifetime Cocaine or Heroin Use at T1</b>				0.58	0.45
Yes	1.23	0.20 (0.27)	0.45		
No	1.00	0.00 (0.00)	.		
<b>Tract-Level Characteristics</b>					
<b>Region</b>				0.25	0.86
Northeast	1.05	0.05 (0.29)	0.86		
Midwest	0.91	-0.10 (0.29)	0.74		
South	0.85	-0.17 (0.28)	0.55		
West	1.00	0.00 (0.00)	.		
<b>Population Density</b>				1.95	0.10
Segment in a CBSA with $\geq$ 1 Million Persons	1.54	0.43 (0.55)	0.44		
Segment in a CBSA with 250,000 to 999,999 Persons	1.00	-0.00 (0.55)	1.00		
Segment in a CBSA with < 250,000 Persons	1.31	0.27 (0.50)	0.59		
Segment Not in a CBSA and Not in a Rural Area	0.41	-0.89 (0.67)	0.18		
Segment Not in a CBSA and in a Rural Area	1.00	0.00 (0.00)	.		
<b>Urbanicity</b>				0.97	0.41
Rural	1.00	0.00 (0.00)	.		
Non-MSA Urban	1.81	0.60 (0.46)	0.20		
MSA < 1 Million	2.35	0.85 (0.51)	0.09		
MSA $\geq$ 1 Million	1.87	0.63 (0.45)	0.17		
<b>Percent Families below Poverty Level</b>	0.38	-0.97 (1.31)	0.46	0.55	0.46
<b>Percent Housing Units Rented</b>	0.23	-1.47 (0.59)	0.01	6.25	0.01
<b>Percent Bachelor, Graduate, or Professional Degrees</b>	2.87	1.05 (0.94)	0.26	1.27	0.26
<b>Median Household Income (in Multiples of \$10,000)</b>	0.73	-0.31 (0.09)	0.00	13.58	0.00

CBSA = core-based statistical area; MSA = metropolitan statistical area; SE = standard error.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.50. Response Propensity Model for Completion of the NSDUH Interview among Main Study Sample Members Aged 12 or Older: 2006 NSDUH Reliability Study**

Completion of the NSDUH Interview Variable	Odds Ratio	Coefficient (SE)	p Value	Wald F	p Value (Wald F)
<b>Intercept</b>	3.31	1.20 (0.11)	0.00		
<b>Age</b>				466.81	0.00
12 to 17	2.34	0.85 (0.03)	0.00		
18 to 25	1.66	0.51 (0.03)	0.00		
26 or Older	1.00	0.00 (0.00)	.		
<b>Gender</b>				56.02	0.00
Male	0.82	-0.20 (0.03)	0.00		
Female	1.00	0.00 (0.00)	.		
<b>Race/Ethnicity</b>				25.71	0.00
White	1.00	0.00 (0.00)	.		
Black or African American	1.14	0.13 (0.06)	0.03		
Other	0.62	-0.48 (0.06)	0.00		
Hispanic or Latino	1.13	0.12 (0.05)	0.01		
<b>Tract-Level Characteristics</b>					
<b>Region</b>				2.95	0.03
Northeast	0.90	-0.10 (0.05)	0.04		
Midwest	1.03	0.03 (0.05)	0.54		
South	0.99	-0.01 (0.05)	0.82		
West	1.00	0.00 (0.00)	.		
<b>Population Density</b>				1.25	0.29
Segment in a CBSA with $\geq 1$ Persons	0.99	-0.01 (0.10)	0.89		
Segment in a CBSA with 250,000 to 999,999 Persons	1.07	0.07 (0.09)	0.48		
Segment in a CBSA with $< 250,000$ Persons	1.11	0.11 (0.09)	0.23		
Segment Not in a CBSA and Not in a Rural Area	1.03	0.03 (0.15)	0.87		
Segment Not in a CBSA and in a Rural Area	1.00	0.00 (0.00)	.		
<b>Urbanicity</b>				1.40	0.24
Rural	1.00	0.00 (0.00)	.		
Non-MSA Urban	1.00	-0.00 (0.07)	0.98		
MSA $< 1$ Million	0.94	-0.06 (0.07)	0.41		
MSA $\geq 1$ Million	0.88	-0.12 (0.07)	0.10		
<b>Percent Families below Poverty Level</b>	1.25	0.22 (0.26)	0.39	0.73	0.39
<b>Percent Housing Units Rented</b>	1.05	0.05 (0.11)	0.68	0.18	0.68
<b>Percent Bachelor, Graduate, or Professional Degrees</b>	0.78	-0.25 (0.17)	0.14	2.24	0.14
<b>Median Household Income (in Multiples of \$10,000)</b>	0.98	-0.02 (0.02)	0.34	0.92	0.34

CBSA = core-based statistical area; MSA = metropolitan statistical area; SE = standard error.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).



**Table 6.51. Dependence of Age at First Use of Cigarettes Consistency and Duration of Recall for Persons Aged 18 or Older: 2006 NSDUH Reliability Study**

Cigarette Prediction Variable	Exact Agreement				Approximate Agreement			
	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)
<b>Intercept</b>	3.43	1.23 (0.44)	0.01		10.02	2.30 (0.56)	0.00	
<b>Years Since First Use of Cigarettes</b>	0.99	-0.01 (0.01)	0.20	0.20	0.98	-0.02 (0.01)	0.02	0.02
<b>Gender</b>				0.09				0.08
Male	0.70	-0.35 (0.21)	0.09		0.62	-0.48 (0.28)	0.08	
Female	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
<b>Race/Ethnicity</b>				0.01				0.15
White	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
Black or African American	0.32	-1.13 (0.38)	0.00		0.47	-0.75 (0.39)	0.05	
Other	0.38	-0.98 (0.57)	0.09		0.58	-0.55 (0.83)	0.51	
Hispanic or Latino	0.86	-0.15 (0.46)	0.74		0.41	-0.88 (0.50)	0.08	
<b>Education</b>				0.05				0.19
< High School	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
High School Graduate	1.60	0.47 (0.33)	0.15		1.48	0.39 (0.43)	0.36	
Some College	1.01	0.01 (0.37)	0.98		0.89	-0.11 (0.46)	0.81	
College Graduate	0.63	-0.46 (0.36)	0.20		0.60	-0.50 (0.46)	0.28	
<b>Same Field Interviewer</b>				0.40				0.71
Yes	0.82	-0.20 (0.23)	0.40		1.12	0.12 (0.31)	0.71	
No	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	

SE = standard error.

NOTE: Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified for each variable. The agreement criterion for approximate agreement was relaxed by 1 year of age.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.52. Dependence of Age at First Use of Alcohol Consistency and Duration of Recall for Persons Aged 18 or Older: 2006 NSDUH Reliability Study**

Alcohol Prediction Variable	Exact Agreement				Approximate Agreement			
	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)
<b>Intercept</b>	0.75	-0.28 (0.39)	0.47		3.96	1.38 (0.40)	0.00	
<b>Years Since First Use of Alcohol</b>	0.99	-0.01 (0.01)	0.03	0.03	0.96	-0.04 (0.01)	0.00	0.00
<b>Gender</b>				0.01				0.21
Male	0.59	-0.53 (0.21)	0.01		0.71	-0.34 (0.27)	0.21	
Female	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
<b>Race/Ethnicity</b>				0.93				0.07
White	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
Black or African American	0.90	-0.10 (0.35)	0.77		1.00	0.00 (0.38)	0.99	
Other	1.32	0.28 (0.46)	0.55		2.63	0.97 (0.58)	0.10	
Hispanic or Latino	0.98	-0.02 (0.39)	0.95		0.45	-0.80 (0.41)	0.05	
<b>Education</b>				0.00				0.00
< High School	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
High School Graduate	2.91	1.07 (0.33)	0.00		2.30	0.83 (0.36)	0.02	
Some College	3.73	1.32 (0.35)	0.00		3.36	1.21 (0.36)	0.00	
College Graduate	3.63	1.29 (0.32)	0.00		4.46	1.50 (0.43)	0.00	
<b>Same Field Interviewer</b>				0.45				0.27
Yes	1.18	0.17 (0.22)	0.45		1.33	0.29 (0.26)	0.27	
No	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	

SE = standard error.

NOTE: Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified for each variable. The agreement criterion for approximate agreement was relaxed by 1 year of age.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.53. Dependence of Age at First Use of Marijuana and Hashish Consistency and Duration of Recall for Persons Aged 18 or Older: 2006 NSDUH Reliability Study**

Marijuana and Hashish Prediction Variable	Exact Agreement				Approximate Agreement			
	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)
<b>Intercept</b>	2.94	1.08 (0.50)	0.03		7.83	2.06 (0.59)	0.00	
<b>Years Since First Use of Marijuana and Hashish</b>	0.98	-0.02 (0.01)	0.03	0.03	0.99	-0.01 (0.02)	0.48	0.48
<b>Gender</b>				0.02				0.16
Male	0.55	-0.59 (0.25)	0.02		0.59	-0.52 (0.37)	0.16	
Female	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
<b>Race/Ethnicity</b>				0.08				0.81
White	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
Black or African American	0.48	-0.74 (0.40)	0.07		0.79	-0.23 (0.67)	0.73	
Other	0.38	-0.97 (0.53)	0.07		0.66	-0.42 (0.73)	0.57	
Hispanic or Latino	1.26	0.23 (0.45)	0.61		1.46	0.38 (0.67)	0.57	
<b>Education</b>				0.97				0.14
< High School	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
High School Graduate	1.13	0.13 (0.46)	0.79		1.57	0.45 (0.56)	0.42	
Some College	1.01	0.01 (0.46)	0.99		1.11	0.10 (0.47)	0.82	
College Graduate	1.14	0.13 (0.41)	0.74		2.77	1.02 (0.50)	0.04	
<b>Same Field Interviewer</b>				0.58				0.06
Yes	0.86	-0.16 (0.28)	0.58		0.48	-0.74 (0.38)	0.06	
No	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	

SE = standard error.

NOTE: Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified for each variable. The agreement criterion for approximate agreement was relaxed by 1 year of age.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.54. Dependence of Age at First Use of Cigarettes Consistency and Duration of Recall for Persons Aged 24 or Older: 2006 NSDUH Reliability Study**

Cigarette Prediction Variable	Exact Agreement				Approximate Agreement			
	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)
<b>Intercept</b>	3.55	1.27 (0.56)	0.02		11.05	2.40 (0.68)	0.00	
<b>Years Since First Use of Cigarettes</b>	0.99	-0.01 (0.01)	0.30	0.30	0.98	-0.02 (0.01)	0.05	0.05
<b>Gender</b>				0.06				0.08
Male	0.67	-0.40 (0.21)	0.06		0.60	-0.52 (0.29)	0.08	
Female	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
<b>Race/Ethnicity</b>				0.01				0.10
White	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
Black or African American	0.30	-1.22 (0.41)	0.00		0.45	-0.79 (0.41)	0.06	
Other	0.31	-1.16 (0.63)	0.06		0.48	-0.72 (0.87)	0.41	
Hispanic or Latino	0.64	-0.45 (0.57)	0.43		0.30	-1.21 (0.59)	0.04	
<b>Education</b>				0.06				0.23
< High School	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
High School Graduate	1.64	0.49 (0.39)	0.20		1.43	0.35 (0.49)	0.47	
Some College	0.97	-0.03 (0.44)	0.94		0.82	-0.20 (0.53)	0.71	
College Graduate	0.62	-0.47 (0.42)	0.27		0.58	-0.55 (0.53)	0.30	
<b>Same Field Interviewer</b>				0.42				0.82
Yes	0.82	-0.20 (0.25)	0.42		1.08	0.08 (0.33)	0.82	
No	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	

SE = standard error.

NOTE: Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified for each variable. The agreement criterion for approximate agreement was relaxed by 1 year of age.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.55. Dependence of Age at First Use of Alcohol Consistency and Duration of Recall for Persons Aged 24 or Older: 2006 NSDUH Reliability Study**

Alcohol Prediction Variable	Exact Agreement				Approximate Agreement			
	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)
<b>Intercept</b>	0.60	-0.51 (0.50)	0.30		4.02	1.39 (0.48)	0.00	
<b>Years Since First Use of Alcohol</b>	0.99	-0.01 (0.01)	0.04	0.04	0.96	-0.04 (0.01)	0.00	0.00
<b>Gender</b>				0.05				0.30
Male	0.63	-0.46 (0.23)	0.05		0.74	-0.30 (0.29)	0.30	
Female	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
<b>Race/Ethnicity</b>				0.67				0.05
White	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
Black or African American	0.80	-0.22 (0.40)	0.58		0.87	-0.13 (0.41)	0.74	
Other	1.70	0.53 (0.49)	0.27		3.28	1.19 (0.61)	0.05	
Hispanic or Latino	1.11	0.10 (0.46)	0.82		0.40	-0.90 (0.46)	0.05	
<b>Education</b>				0.00				0.00
< High School	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
High School Graduate	3.72	1.31 (0.40)	0.00		2.35	0.85 (0.40)	0.03	
Some College	5.30	1.67 (0.42)	0.00		3.81	1.34 (0.40)	0.00	
College Graduate	4.41	1.48 (0.38)	0.00		4.55	1.51 (0.45)	0.00	
<b>Same Field Interviewer</b>				0.46				0.35
Yes	1.20	0.18 (0.24)	0.46		1.30	0.26 (0.28)	0.35	
No	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	

SE = standard error.

NOTE: Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified for each variable. The agreement criterion for approximate agreement was relaxed by 1 year of age.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.56. Dependence of Age at First Use of Marijuana and Hashish Consistency and Duration of Recall for Persons Aged 24 or Older: 2006 NSDUH Reliability Study**

Marijuana and Hashish Prediction Variable	Exact Agreement				Approximate Agreement			
	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)	Odds Ratio	Coefficient (SE)	<i>p</i> Value	<i>p</i> Value (Wald F)
<b>Intercept</b>	3.48	1.25 (0.68)	0.07		5.19	1.65 (0.73)	0.02	
<b>Years Since First Use of Marijuana and Hashish</b>	0.98	-0.02 (0.01)	0.09	0.09	1.00	0.00 (0.02)	0.90	0.90
<b>Gender</b>				0.04				0.13
Male	0.56	-0.58 (0.28)	0.04		0.55	-0.59 (0.39)	0.13	
Female	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
<b>Race/Ethnicity</b>				0.09				0.63
White	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
Black or African American	0.51	-0.68 (0.48)	0.16		0.82	-0.20 (0.76)	0.79	
Other	0.38	-0.96 (0.58)	0.10		0.68	-0.39 (0.74)	0.60	
Hispanic or Latino	1.96	0.67 (0.52)	0.20		2.31	0.84 (0.79)	0.29	
<b>Education</b>				1.00				0.13
< High School	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	
High School Graduate	0.93	-0.07 (0.57)	0.90		1.54	0.43 (0.61)	0.48	
Some College	0.88	-0.12 (0.58)	0.83		1.13	0.12 (0.52)	0.82	
College Graduate	0.97	-0.03 (0.51)	0.95		2.99	1.09 (0.54)	0.04	
<b>Same Field Interviewer</b>				0.47				0.08
Yes	0.80	-0.23 (0.31)	0.47		0.48	-0.74 (0.42)	0.08	
No	1.00	0.00 (0.00)	.		1.00	0.00 (0.00)	.	

SE = standard error.

NOTE: Approximate agreement is when responses at both interviews are the same or nearly the same, where "nearly the same" is specified for each variable. The agreement criterion for approximate agreement was relaxed by 1 year of age.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

**Table 6.57. Predicted Marginals for Proportion of Agreement between T1 and T2: 2006 NSDUH Reliability Study**

Substance	Agreement Criterion	Years Since First Use	Predicted Marginal (SE)	
			Model 1	Model 2
Cigarettes	Exact	5	0.66 (0.03)	0.67 (0.03)
		10	0.65 (0.03)	0.66 (0.03)
		15	0.64 (0.04)	0.65 (0.03)
	Approximate	5	0.85 (0.02)	0.85 (0.02)
		10	0.84 (0.02)	0.84 (0.02)
		15	0.83 (0.02)	0.83 (0.02)
Alcohol	Exact	5	0.64 (0.03)	0.63 (0.04)
		10	0.62 (0.03)	0.62 (0.03)
		15	0.61 (0.03)	0.60 (0.03)
	Approximate	5	0.90 (0.02)	0.89 (0.02)
		10	0.88 (0.02)	0.87 (0.02)
		15	0.86 (0.02)	0.85 (0.02)
Marijuana	Exact	5	0.62 (0.04)	0.62 (0.05)
		10	0.60 (0.04)	0.59 (0.04)
		15	0.57 (0.03)	0.57 (0.04)
	Approximate	5	0.82 (0.04)	0.83 (0.05)
		10	0.82 (0.04)	0.82 (0.04)
		15	0.81 (0.03)	0.81 (0.03)

NOTE: Model 1 was fitted to all ages (i.e., those aged 12 years old or older) who had nonmissing age at first use data both at T1 and T2, without education data. Model 2 was fitted to those aged 18 years old or older who had nonmissing age at first use data both at T1 and T2, with education level as a covariate.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.58. Estimated True Prevalence of Past Year Marijuana Use and False Negative Rates, by Gender, among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Gender	Estimated True Prevalence Rates	Estimated False Negative Rates
Male	0.16	0.20
Female	0.08	0.20

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.59. Estimated True Prevalence of Past Year Marijuana Use and False Negative Rates, by Race/Ethnicity, among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Race/Ethnicity	Estimated True Prevalence Rates	Estimated False Negative Rates
White, Non-Hispanic	0.13	0.12
Black, Non-Hispanic	0.30	0.45
Other, Non-Hispanic	0.34	0.40
Hispanic or Latino	0.25	0.16

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.60. Estimated True Prevalence of Past Year Marijuana Use and False Negative Rates, by Age Group, among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Age Group	Estimated True Prevalence Rates	Estimated False Negative Rates
12 to 17	0.17	0.30
18 to 25	0.42	0.12
26 or Older	0.12	0.23

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.61. Estimated True Prevalence of Past Year Marijuana Use and False Negative Rates, by the U.S. Born Indicator, among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

U.S. Born	Estimated True Prevalence Rates	Estimated False Negative Rates
Yes	0.17	0.19
No	0.06	0.64

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

**Table 6.62. Estimated True Prevalence of Past Year Marijuana Use and False Negative Rates, by Metropolitan Status, among Persons Aged 12 or Older: 2006 NSDUH Reliability Study**

Metropolitan Status	Estimated True Prevalence Rates	Estimated False Negative Rates
Yes	0.17	0.21
No	0.10	0.16

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).



## 7. Reliability Findings for Same versus Different Interviewer Substudy

The rationale and design for the same versus different interviewer substudy are provided in Section 3.1.2 in Chapter 3. It was of interest to know whether a different interviewer from the one who conducted the initial interview (T1) would have any biasing influence on responses to the follow-up interview (T2). Because most of the information collected in the National Survey on Drug Use and Health (NSDUH) is obtained through the audio computer-assisted self-interviewing (ACASI) portion of the instrument, it was hypothesized that the field interviewer (FI) would have little or no effect on the reliability of survey responses. However, the FI does interact with the respondent in several ways:

- The FI explains the study and its importance in order to motivate the respondent to participate.
- The FI provides informed consent information, including the statement of confidentiality of responses in order to motivate the respondent to answer all questions truthfully.
- The FI answers respondent questions before and during the interview administration.

Additionally, small but somewhat persistent interviewer experience effects on substance use prevalence estimates have been observed previously in the main study. Thus, it may still be possible for some FI effect to exist.

Therefore, statistical testing of the hypothesis of "no difference in kappa values between the same FI (individuals who had both their interviews administered by the same FI) group and the different FI group" was conducted for selected outcome variables. As already noted in this report, comparison of kappa values between groups where prevalence rates are not similar may be misleading because of the sensitivity of the kappa measure to prevalence rates. However, because there was a random assignment of the two study conditions to clusters of sample housing units, it was assumed that there was no inherent difference in prevalence rates among the clusters assigned to the "same FI" or "different FI" conditions.

Table 7.1 provides the results of the significance testing. Kappa values were the same or had only small differences between the "same FI" and "different FI" groups. Indeed, only one  $p$  value was significant at the 0.05 level. This is reassuring for two reasons. First, it indicates that the administration by the same FI or different FI did not affect the reliability of survey responses. Second, observing kappa values so similar between the two groups indicates that the precision of the estimated kappas is high.

**Table 7.1. Impact of the Administration of Both Interviews by the Same Field Interviewer (FI) or by Different FIs on the Reliability of Selected Measures: 2006 NSDUH Reliability Study**

Variable	Same FI			Different FI			Difference Kappa	t Statistic	p Value
	N	Kappa	SE	n	Kappa	SE			
<b>Past Year Marijuana Use</b>	2,063	0.82	0.04	1,065	0.82	0.05	0.00	-0.03	0.97
<b>K6 Score<sup>1</sup></b>									
Exact Agreement	1,410	0.21	0.03	718	0.21	0.03	0.00	0.02	0.98
Approximate Agreement	1,410	0.64	0.04	718	0.64	0.05	0.00	-0.05	0.96
<b>Dependence or Abuse – Illicit Drugs or Alcohol</b>	981	0.70	0.07	541	0.62	0.08	0.07	0.73	0.47
<b>Age</b>									
12 to 17	101	0.34	0.14	71	0.63	0.17	0.29	-1.30	0.19
18 to 25	419	0.53	0.09	226	0.49	0.13	0.04	0.29	0.77
26 or Older	461	0.74	0.08	244	0.62	0.11	0.12	0.84	0.40
<b>Days Missed School or Work<sup>2</sup></b>									
Exact Agreement	1,541	0.47	0.05	804	0.43	0.05	0.04	0.50	0.62
Approximate Agreement	1,541	0.56	0.14	804	0.39	0.09	0.17	1.00	0.32
<b>Annual Family Income<sup>3</sup></b>									
Exact Agreement	1,810	0.77	0.04	914	0.75	0.05	0.01	0.21	0.83
Approximate Agreement	1,810	1.00	0.00	914	0.99	0.00	0.01	1.56	0.12
<b>Industry Code</b>									
4-Digit Code	1,288	0.73	0.03	658	0.63	0.04	0.11	1.96	0.05
2-Digit Code	1,288	0.84	0.03	658	0.80	0.03	0.04	0.96	0.34
<b>Occupation Code</b>									
4-Digit Code	1,288	0.77	0.03	658	0.68	0.04	0.08	1.75	0.08
2-Digit Code	1,288	0.85	0.03	658	0.81	0.03	0.04	1.04	0.30

SE = standard error.

<sup>1</sup> The agreement criterion for approximate agreement was relaxed by three points.

<sup>2</sup> This variable was created based on the number of days the respondent missed school or work due to sickness or injury in the past 30 days. The agreement criterion for approximate agreement was relaxed by 3 days.

<sup>3</sup> Levels for this annual family income variable are "less than \$20,000," "\$20,000-\$49,999," "\$50,000-\$74,999," and "\$75,000 or more." The agreement criterion for approximate agreement was relaxed by one category.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (n = 3,136).

## 8. Impact of Time between Interviews on Reliability

One of the main assumptions in the initial interview (T1) and reinterview (T2) methodology for the 2006 National Survey on Drug Use and Health (NSDUH) Reliability Study was that respondents would not make a conscious effort to recall and repeat their earlier responses at the second interview. As more time passed since the first interview, this became less of a concern because respondents would be more likely to forget their responses at the first interview. Therefore, it was of interest to test whether there was a statistically significant effect of the number of days between the two interviews. (See Section 5.4 for the distribution of the number of days between interviews.)

A test was conducted of the effect of the number of days between the interviews by dividing the sample into two nearly equal groups: (a) those who had their second interview less than 9 days after the first, and (b) those whose two interviews had 9 or more days between them (the weighted median of the time between interviews was 8 days). The kappa values for selected outcome variables were compared, and the differences tested for significance. Similar to the testing in Chapter 6, comparing the kappa values between the two groups is legitimate because the number of days between interviews is independent of the prevalence rates.

Table 8.1 provides the results of the significance testing. Kappa values between the "less than 9 days" and "9 days or more" between interviews were the same or had only small differences, with the exception of one outcome, "dependence or abuse – any illicit drug or alcohol," where the  $p$  value was 0.00. Interestingly, and unexpectedly, the kappa for those who had their interviews less than 9 days apart was 0.55, whereas those whose second interview was further separated from their first interview had a kappa of 0.80. Note that multiple testing often results in at least one significant  $p$  value because of chance alone. The only significant difference here is in the counterintuitive direction, so this may not be a true difference, but rather an artifact of random sampling.

Additional information on the distribution of responses to the follow-up questions by the number of days between interviews is presented in Appendix C.

**Table 8.1. Impact of the Time between Interviews on the Reliability of Selected Measures: 2006 NSDUH Reliability Study**

Variable	Time between Interviews: Less Than 9 Days			Time between Interviews: 9 Days or More			Difference Kappa	t Statistic	p Value
	n	Kappa	SE	n	Kappa	SE			
<b>Lifetime Use</b>									
Cigarettes	1,904	0.92	0.02	1,232	0.92	0.03	0.00	-0.11	0.91
Alcohol	1,903	0.85	0.03	1,232	0.80	0.07	0.05	0.65	0.51
Marijuana	1,902	0.93	0.02	1,230	0.94	0.02	-0.01	-0.35	0.73
<b>Past Year Use</b>									
Cigarettes	1,899	0.93	0.02	1,230	0.92	0.02	0.01	0.44	0.66
Alcohol	1,901	0.90	0.02	1,229	0.89	0.04	0.01	0.34	0.73
Marijuana	1,900	0.78	0.05	1,228	0.86	0.04	-0.07	-1.20	0.23
<b>Age at First Use – Exact Agreement</b>									
Cigarettes	1,050	0.60	0.04	701	0.57	0.05	0.03	0.45	0.65
Alcohol	1,312	0.50	0.03	869	0.52	0.05	-0.01	-0.22	0.82
Marijuana	704	0.51	0.05	456	0.48	0.05	0.03	0.45	0.65
<b>Age at First Use – Approximate Agreement<sup>1</sup></b>									
Cigarettes	1,050	0.72	0.04	701	0.70	0.06	0.01	0.21	0.83
Alcohol	1,312	0.68	0.04	869	0.68	0.06	0.00	0.06	0.95
Marijuana	704	0.77	0.05	456	0.70	0.07	0.07	0.77	0.44
<b>K6 Score<sup>2</sup></b>									
Exact Agreement	1,286	0.21	0.03	842	0.21	0.04	0.00	-0.08	0.94
Approximate Agreement	1,286	0.63	0.04	842	0.65	0.06	-0.02	-0.23	0.82
<b>Dependence or Abuse – Illicit Drugs or Alcohol</b>	911	0.55	0.06	611	0.80	0.06	-0.26	-3.02	0.00

SE = standard error.

<sup>1</sup> The agreement criterion for approximate agreement was relaxed by 1 day.

<sup>2</sup> The agreement criterion for approximate agreement was relaxed by three points.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study ( $n = 3,136$ ).

## 9. Recommendations for Future Reliability Studies

The Reliability Study described in this report provides useful information about the reliability of survey responses, and this report provides general procedural guidelines for future studies of reliability. Because of the length and sensitivity of the National Survey on Drug Use and Health (NSDUH) interview, providing an adequate monetary incentive is probably paramount to achieving a high response rate and to getting the respondent to agree to the reinterview within the time window desired. Shorter interviews with less sensitive content and a less stringent reinterview time window would likely not require as great a monetary incentive as was used in this study.

The substudy comparing respondents interviewed by the same interviewer for both interviews with respondents interviewed by different interviewers showed there was no consistent evidence of the presence of an interviewer effect for the reinterview. Because the major part of the NSDUH interview uses audio computer-assisted self-interviewing (ACASI), any potential interviewer effect is likely diminished. Surveys with more interviewer-administered questions may be more prone to showing interviewer effects. The extent to which that is the case has yet to be determined.

Although there were no consistent significant differences in reliability because of the length of time between the initial interview and the reinterview, the follow-up questions (see Appendix C) generally showed that respondents with less time between the T1 and T2 interviews remembered the questions and their initial answers better than did respondents with more time between interviews. Thus, it was felt that the more time that can be allowed between interviews, the better the conditions for testing reliability of questions should be. However, this does have to be considered in relation to the reference period for the questions. In the NSDUH case, those questions that ask about behaviors that occurred in the past 30 days are subject to showing less reliability simply because the reference period is so short that the time between the initial and follow-up interviews could legitimately result in the initial and follow-up interview responses being different because of the time period covered by the reference period. Systematically varying the interval between interviews to examine effects on reliability could be a useful manipulation under certain conditions.

Because of the potential for context effects in questionnaire design, readministering the same questionnaire at both the initial and reinterview times, as in the 2006 Reliability Study, is recommended. Subsetting the questionnaire to a set of what the researcher considers the "key" questionnaire items has the potential for introducing unexpected context effects that could mask the reliability of the questions tested.

The analysis of the NSDUH questions in this study showed that questions about factual personal events were more reliable than questions that asked for a respondent's opinion or intentions or questions that addressed issues that carried social stigma. Thus, a test/retest protocol to assess the reliability of opinion or feeling questions, questions about a person's intentions, questions about issues that carry a social stigma, or questions of personal activities

that can change rapidly in a matter of days may yield lower reliability responses. Although these results are not surprising, not many studies in the literature have similar findings. Comparing the reliability of questions on behaviors or attitudes that vary in social desirability as perceived by the researcher could yield insights into the relative extent to which they are stigmatized by the population being surveyed. Analyses involving Hui-Walter modeling may also be able to shed light on these issues.

# References

- Agresti, A. (2002). *Categorical data analysis* (No. 359, Wiley Series in Probability and Statistics, 2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc., Wiley Interscience.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)* (4th ed.). Washington, DC: Author.
- Bowman, K., Caviness, L., Coffey, L. E., Cunningham, D., Jones, B., Liu, S., Moffat, A., Myers, S., Penne, M., & Piper, L. (2004, July). *2003 National Survey on Drug Use and Health (NSDUH): Data collection final report* (prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-98-9008, RTI/7190-2003 NSDUH). Research Triangle Park, NC: RTI International.
- Brick, J. M., Rizzo, L., & Wernimont, J. (1997, June). *Reinterview results for the school safety and discipline and school readiness components* (NCES 97339). Washington, DC: U.S. Department of Education, National Center for Education Statistics. [Available as a PDF at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=97339>]
- Chen, P., Dai, L., Gordek, H., Laufenberg, J., Liu, B., Sathe, N. S., & Westlake, M. (2008, January). Person-level sampling weight calibration. In *2006 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Phase I, Deliverable No. 39, RTI/0209009.274.002). Research Triangle Park, NC: RTI International.
- Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the 1979 American Statistical Association, Survey Research Methods Section, Washington, DC* (pp. 401-406). Washington, DC: American Statistical Association. [Available as a PDF at <http://www.amstat.org/sections/SRMS/proceedings/>]
- Citro, C. F., & Cohen, M. L. (1985). *The bicentennial census: New directions for methodology in 1990*. Washington, DC: National Academy Press.
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley & Sons, Inc.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213-220.
- Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

Feder, M. (2006). Variance estimation of the survey-weighted kappa measure of agreement. In *Proceedings of the 2006 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Seattle, WA* (CD-ROM, pp. 3002-3007). Alexandria, VA: American Statistical Association. [Available at <http://www.amstat.org/sections/SRMS/proceedings/>]

Federal Committee on Statistical Methodology. (2001, July). *Measuring and reporting sources of error in surveys* (Statistical Policy Working Paper 31). Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office. [Available as a PDF at <http://www.fcsm.gov/reports/>]

Fleiss, J. L. (1975). Measuring agreement between two judges on the presence or absence of a trait. *Biometrics*, *31*, 651-659.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*, 613-619.

Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, *72*, 323-327.

Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 598-603). Alexandria, VA: American Statistical Association. [Available as a PDF at <http://www.amstat.org/sections/SRMS/proceedings/>]

Forsman, G., & Schreiner, I. (1991). The design and analysis of reinterview: An overview. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 279-302). New York: John Wiley & Sons, Inc.

Harrison, L. D., Martin, S. S., Enev, T., & Harrington, D. (2007). *Comparing drug testing and self-report of drug use among youths and young adults in the general population* (DHHS Publication No. SMA 07-4249, Methodology Series M-7). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Hui, S. L., & Walter, S. D. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, *36*, 167-171.

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

Kroutil, L. A., & Handley, W. (2008, January 8). General principles and procedures for editing drug use data in the 2006 NSDUH computer-assisted interview. In *2006 National Survey on Drug Use and Health: Methodological resource book* (Section 10, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009.273). Research Triangle Park, NC: RTI International.



Kulka, R. A., Schlenger, W. E., Fairbank, J. A., Hough, R. L., Jordan, B. K., Marmar, C. R., & Weiss, D. S. (1990). *Trauma and the Vietnam War generation: Report of findings from the National Vietnam Veterans Readjustment Study*. New York: Brunner/Mazel.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

Morton, K. B., Chromy, J. R., Hunter, S. R., & Martin, P. C. (2007, January; revised February 2008). Sample design report. In *2006 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Phase II, Deliverable No. 8, RTI/0209009.230.004). Research Triangle Park, NC: RTI International.

Office of Applied Studies. (2007). *Results from the 2006 National Survey on Drug Use and Health: National findings* (DHHS Publication No. SMA 07-4293, NSDUH Series H-32). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Management and Budget. (2006, September). *Standards and guidelines for statistical surveys*. Washington, DC: Author. [Available as a PDF at [http://www.whitehouse.gov/omb/assets/omb/inforeg/statpolicy/standards\\_stat\\_surveys.pdf](http://www.whitehouse.gov/omb/assets/omb/inforeg/statpolicy/standards_stat_surveys.pdf)]

Piper, L., Meyer, M., & Snodgrass, J. (2006, July). *2005 National Survey on Drug Use and Health: Reliability study pretest final report* (prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 31, RTI 0209009.124). Research Triangle Park, NC: RTI International.

Pritzker, L., & Hanson, R. (1962). Measurement errors in the 1960 Census of Population. In *Proceedings of the Social Statistics Section, American Statistical Association* (pp. 80-89). Washington, DC: American Statistical Association.

RTI International. (2004). *SUDAAN<sup>®</sup> language manual, Release 9.0*. Research Triangle Park, NC: Author.

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*, 7, 301-317.

Shrout, P. E., Spitzer, R. L., & Fleiss, J. L. (1987). Quantification of agreement in psychiatric diagnosis revisited. *Archives of General Psychiatry*, 44, 172-177.

Sinclair, M. D., & Gastwirth, J. L. (1996). On procedures for evaluating the effectiveness of reinterview survey methods: Application to labor force data. *Journal of the American Statistical Association*, 91, 961-969.

Singer, P., & Ennis, S. R. (2003, September 24.). *Census 2000 Content Reinterview Survey: Accuracy of data for selected population and housing characteristics as measured by re-interview* (Census 2000 Evaluation B.5). Washington, DC: U.S. Census Bureau. [Available at [http://www.census.gov/pred/www/eval\\_top\\_rpts.htm](http://www.census.gov/pred/www/eval_top_rpts.htm)]

Taeuber, C., & Hansen, M. H. (1964). A preliminary evaluation of the 1960 Censuses of Population and Housing. *Demography*, *1*(1), 1-14.

Thompson, W. D., & Walter, S. D. (1988). A reappraisal of the kappa coefficient. *Journal of Clinical Epidemiology*, *41*, 949-958.

Vermunt, J. K. (1997). *LEM 1.0: A general program for the analysis of categorical data*. Tilburg, The Netherlands: Tilburg University.

Wright, D., Sathe, N., & Spagnola, K. (2007). *State estimates of substance use from the 2004-2005 National Surveys on Drug Use and Health* (DHHS Publication No. SMA 07-4235, NSDUH Series H-31). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

# Appendix A: Description of NSDUH

This appendix describes the sample design, data collection methodology, and data processing for the main study of the 2006 National Survey on Drug Use and Health<sup>8</sup> (NSDUH). Special features of Spanish interviews and editing and weighting procedures for the NSDUH Reliability Study conducted in 2006 are discussed.

## A.1. Sample Design

The 2006 NSDUH is part of a coordinated 5-year sample design providing estimates for all 50 States plus the District of Columbia for the years 2005 through 2009. The respondent universe is the civilian, noninstitutionalized population aged 12 years old or older residing within the United States. The survey includes persons living in noninstitutionalized group quarters (e.g., shelters, rooming/boarded houses, college dormitories, migratory workers' camps, halfway houses), and civilians living on military bases. Persons excluded from the survey include persons with no fixed household address (e.g., homeless and/or transient persons not in shelters), active-duty military personnel, and residents of institutional group quarters, such as correctional facilities, nursing homes, mental institutions, and long-term hospitals.

Although there is no planned overlap with the 1999 through 2004 samples, a coordinated design for 2005 through 2009 facilitates 50 percent overlap in second-stage units (area segments) within each successive 2-year period from 2005 through 2009. Because the 2005 design enables estimates to be developed by State in all 50 States plus the District of Columbia, States may be viewed as the first level of stratification as well as a reporting variable.

For the 50-State design, 8 States were designated as large sample States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) with target sample sizes of 3,600. In 2006, sample sizes in these States ranged from 3,512 to 3,671. For the remaining 42 States and the District of Columbia, the target sample size was 900. Sample sizes in these States ranged from 862 to 1,000 in 2006. This approach ensures there is sufficient sample in every State to support small area estimation (SAE)<sup>9</sup> while at the same time maintaining efficiency for national estimates.

States were first stratified into a total of 900 State sampling (SS) regions (48 regions in each large sample State and 12 regions in each small sample State). These regions were contiguous geographic areas designed to yield the same number of interviews on average.<sup>10</sup> Unlike the 1999 through 2001 NHSDAs and the 2002 through 2004 NSDUHs in which the first-stage sampling units were clusters of census blocks called area segments, the first stage of

---

<sup>8</sup> Prior to 2002, the survey was known as the National Household Survey on Drug Abuse (NHSDA).

<sup>9</sup> SAE is a hierarchical Bayes modeling technique used to make State-level estimates for approximately 20 substance use-related measures. For more details, see the *State Estimates of Substance Use from the 2004-2005 National Surveys on Drug Use and Health* (Wright, Sathe, & Spagnola, 2007).

<sup>10</sup> Areas were defined using 2000 census geography. Dwelling units (DUs) and population counts were obtained from the 2000 census data supplemented with revised population counts from Claritas (<http://cluster1.claritas.com/claritas/Default.jsp>).

selection for the 2005 through 2009 NSDUHs was census tracts.<sup>11</sup> This stage was included to contain sample segments within a single census tract to the extent possible.<sup>12</sup>

For each SS region, 48 census tracts were selected with probability proportional to size. Within sampled census tracts, adjacent census blocks were combined to form the second-stage sampling units or area segments. One area segment was selected within each sampled census tract with probability proportional to population size to support the 5-year sample and any supplemental studies that the Substance Abuse and Mental Health Services Administration (SAMHSA) may choose to field.<sup>13</sup> Of these segments, 24 were designated for the coordinated 5-year sample, and 24 were designated as "reserve" segments. Eight sample segments per SS region were fielded during the 2006 survey year.

These sampled segments were allocated equally into four separate samples, one for each 3-month period (calendar quarter) during the year. That is, a sample was selected from two segments in each calendar quarter so that the survey was essentially continuous in the field. In each of the area segments, a listing of all addresses was made, from which a national sample of 182,459 addresses was selected. Of the selected addresses, 151,288 were determined to be eligible sample units. In these sample units (which can be either households or units within group quarters), sample persons were randomly selected using an automated screening procedure programmed in a handheld computer carried by the interviewers. The number of sample units completing the screening was 137,057. Youths aged 12 to 17 years and young adults aged 18 to 25 years were oversampled at this stage. Because of the large sample size, there was no need to oversample racial/ethnic groups, as was done on surveys prior to 1999. Nationwide, 85,034 persons were selected. Consistent with previous surveys in this series, the final respondent sample of 67,802 persons was representative of the U.S. general population (since 1991, the civilian, noninstitutionalized population) aged 12 or older. In addition, State samples were representative of their respective State populations. More detailed information on the disposition of the national screening and interview sample can be found in Appendix B of the 2006 national results report (Office of Applied Studies [OAS], 2007).

The survey covers residents of households (living in houses/townhouses, apartments, condominiums, etc.), persons in noninstitutional group quarters (e.g., shelters, rooming/boarding houses, college dormitories, migratory workers' camps, halfway houses), and civilians living on military bases. Although the survey covers residents of these types of units (they are given a nonzero probability of selection), the sample sizes of most specific groups are too small to provide separate estimates.

More information on the sample design can be found in the 2006 NSDUH sample design report by Morton et al. (2007-2008).

---

<sup>11</sup> Census tracts are relatively permanent statistical subdivisions of counties and provide a stable set of geographic units across decennial census periods.

<sup>12</sup> Some census tracts had to be aggregated in order to meet the minimum DU requirement of 150 DUs in urban areas and 100 DUs in rural areas.

<sup>13</sup> For more details on the 5-year sample, see the 2006 sample design report in the *2006 NSDUH Methodological Resource Book* (Morton, Chromy, Hunter, & Martin, 2007-2008).

## A.2. Data Collection Methodology

The data collection method used in NSDUH involves in-person interviews with sample persons, incorporating procedures that would be likely to increase respondents' cooperation and willingness to report honestly about their illicit drug use behavior. Confidentiality is stressed in all written and oral communications with potential respondents. Respondents' names are not collected with the data, and computer-assisted interviewing (CAI) methods are used to provide a private and confidential setting to complete the interview.

Introductory letters are sent to sampled addresses, followed by an interviewer visit. A 5-minute screening procedure using a handheld computer involves listing all household members along with their basic demographic data. The computer uses the demographic data in a preprogrammed selection algorithm to select zero to two sample persons, depending on the composition of the household. This selection process is designed to provide the necessary sample sizes for the specified population age groupings. In areas where a third or more of the households contain Spanish-speaking residents, the initial introductory letters written in English are mailed with a Spanish version on the back. All interviewers carry copies of this letter in Spanish. If the interviewer is not certified bilingual, he or she will use preprinted Spanish cards to attempt to find someone in the household who speaks English and who can serve as the screening respondent or who can translate for the screening respondent. If no one is available, the interviewer will schedule a time when a Spanish-speaking interviewer can come to the address. In households where a language other than Spanish is encountered, another language card is used to attempt to find someone who speaks English to complete the screening.

The NSDUH interview is available in English and Spanish, and both versions have the same content. If the sample person prefers to complete the interview in Spanish, a certified bilingual interviewer is sent to the address to conduct the interview. Because the interview is not translated into any other language, if a sample person does not speak English or Spanish, the interview is not conducted. As noted in Section 3.3, sample persons who chose to complete the interview in Spanish were not eligible to participate in the Reliability Study.

Interviewers attempt to conduct the NSDUH interview immediately with each selected person in the household. The interviewer requests the selected respondent to identify a private area in the home to conduct the interview away from other household members. The interview averages about an hour in length and includes a combination of CAPI (computer-assisted personal interviewing, in which the interviewer reads the questions) and ACASI (audio computer-assisted self-interviewing).

The NSDUH interview consists of a core section and noncore (i.e., supplemental) sections. A core set of questions critical for basic trend measurement of prevalence estimates remains in the survey every year and comprises the first part of the interview. Noncore questions, or modules, that can be revised, dropped, or added from year to year make up the remainder of the interview. The core consists of initial demographic items (which are interviewer-administered) and self-administered questions pertaining to the use of tobacco, alcohol, marijuana, cocaine, crack cocaine, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives. Topics in the remaining noncore self-administered sections include (but are not limited to) injection drug use, perceived risks of substance use, substance

dependence or abuse, arrests, treatment for substance use problems, pregnancy and health care issues, and mental health issues. Noncore demographic questions (which are interviewer-administered and follow the ACASI questions) address such topics as immigration, current school enrollment, employment and workplace issues, health insurance coverage, and income. It should be noted that some of the noncore portions of the interview have remained in the survey, relatively unchanged, from year to year (e.g., current health insurance coverage, employment).

Thus, the interview begins in CAPI mode with the field interviewer (FI) reading the questions from the computer screen and entering the respondent's replies into the computer. The interview then transitions to the ACASI mode for the sensitive questions. In this mode, the respondent can read the questions silently on the computer screen and/or listen to the questions read through headphones and enter his or her responses directly into the computer. At the conclusion of the ACASI section, the interview returns to the CAPI mode with the interviewer completing the questionnaire. Each respondent who completes a full interview is given a \$30 cash payment as a token of appreciation for his or her time.

No personal identifying information is captured in the CAI record for the respondent. Interviewers transmit the completed interview data to RTI in Research Triangle Park, North Carolina, via home telephone lines.

### **A.3. Data Processing**

Computers at RTI direct the information to a raw data file (i.e., in which no logical editing or statistical imputation has been done) that consists of one record for each completed interview. Cases are retained only if respondents provided data on lifetime use of cigarettes and at least nine other substances in the core section of the questionnaire. Some editing and consistency checks are done by the CAI program during the interview; however, the output from the CAI is considered the raw data. For the Reliability Study, there were instances where using the strictly raw data was not meaningful. Thus, there was some "light editing" done to some of the data to make the reliability analyses meaningful. The light editing included assignment of numerical codes to certain classes of responses, collapsing all missing and unknown categories to a SAS<sup>®</sup> missing code, and similar types of consistency edits. Also, for questions like the substance use gate questions, if the respondent either refused or responded "don't know," follow-up probing questions were asked. This series of questions was used to determine the outcome variable, and the Reliability Study analyzed this outcome variable instead of the individual component variables. None of the additional, more complex edits and consistency checks nor statistical imputation applied to the main NSDUH data was applied to the Reliability Study data for this analysis. See Section 3.8.1 for more details.

For a description of data coding and logical editing procedures and statistical imputation procedures used to construct datasets used in other analyses of NSDUH data, see the section on data processing in Appendix A in the 2006 national results report (OAS, 2007) or the section on general editing in the *2006 Methodological Resource Book* (Kroutil & Handley, 2008).

Analyses conducted for this report used analysis weights developed for the 2006 NSDUH. The general approach to developing and calibrating analysis weights involved developing design-based weights,  $d_k$ , as the product of the inverse of the selection probabilities at

each selection stage. Similar to the 2005 NSDUH, the 2006 NSDUH used a four-stage sample selection scheme in which an extra selection stage of census tracts was added before the selection of a segment. Thus, the design-based weights,  $d_k$ , for the 2006 NSDUH incorporated the extra layer of sampling selection to reflect the sample design change. Adjustment factors,  $a_k(\lambda)$ , then were applied to the design-based weights to adjust for nonresponse, to poststratify to known population control totals, and to control for extreme weights when necessary. In view of the importance of State-level estimates with the 50-State design, it was necessary to control for a much larger number of known population totals. Several other modifications to the general weight adjustment strategy that had been used in past surveys also were implemented for the first time beginning with the 1999 CAI sample.

Weight adjustments were based on a generalization of Deville and Särndal's (1992) logit model. This generalized exponential model (GEM) (Folsom & Singh, 2000) incorporates unit-specific bounds  $(\ell_k, u_k)$ ,  $k \in S$ , for the adjustment factor  $a_k(\lambda)$  as follows:

$$a_k(\lambda) = \frac{\ell_k(u_k - c_k) + u_k(c_k - \ell_k) \exp(A_k x_k' \lambda)}{(u_k - c_k) + (c_k - \ell_k) \exp(A_k x_k' \lambda)},$$

where  $c_k$  are prespecified centering constants, such that  $\ell_k < c_k < u_k$  and  $A_k = (u_k - \ell_k) / (u_k - c_k)(c_k - \ell_k)$ . The variables  $\ell_k$ ,  $c_k$ , and  $u_k$  are user-specified bounds, and  $\lambda$  is the column vector of  $p$  model parameters corresponding to the  $p$  covariates  $x$ . The  $\lambda$ -parameters are estimated by solving

$$\sum_s x_k d_k a_k(\lambda) - \tilde{T}_x = 0,$$

where  $\tilde{T}_x$  denotes control totals that could be either nonrandom, as is generally the case with poststratification, or random, as is generally the case for nonresponse adjustment.

The final weights,  $w_k = d_k a_k(\lambda)$ , minimize the distance function  $\Delta(w, d)$  defined as

$$\Delta(w, d) = \sum_{k \in S} \frac{d_k}{A_k} \left\{ (a_k - \ell_k) \log \frac{a_k - \ell_k}{c_k - \ell_k} + (u_k - a_k) \log \frac{u_k - a_k}{u_k - c_k} \right\}.$$

This general approach was used at several stages of the weight adjustment process, including (1) adjustment of household weights for nonresponse at the screener level, (2) poststratification of household weights to meet population controls for various demographic groups by State, (3) adjustment of household weights for extremes, (4) poststratification of selected person weights, (5) adjustment of responding person weights for nonresponse at the questionnaire level, (6) poststratification of responding person weights, and (7) adjustment of responding person weights for extremes.

Every effort was made to include as many relevant State-specific covariates (typically defined by demographic domains within States) as possible in the multivariate models used to calibrate the weights (nonresponse adjustment and poststratification steps). Because further subdivision of State samples by demographic covariates often produced small cell sample sizes, it was not possible to retain all State-specific covariates (even after meaningful collapsing of

covariate categories) and still estimate the necessary model parameters with reasonable precision. Therefore, a hierarchical structure was used in grouping States with covariates defined at the national level, at the census division level within the Nation, at the State group within the census division, and, whenever possible, at the State level. In every case, the controls for the total population within a State and the five age groups (12 to 17, 18 to 25, 26 to 34, 35 to 49, 50 or older) within a State were maintained except that, in the last step of poststratification of person weights, six age groups (12 to 17, 18 to 25, 26 to 34, 35 to 49, 50 to 64, 65 or older) were used. Census control totals by age, race, gender, and Hispanicity were required for the civilian, noninstitutionalized population of each State. Beginning with the 2002 NSDUH, the Population Estimates Branch of the U.S. Census Bureau has produced the necessary population estimates in response to a special request based on the 2000 census.

Consistent with the surveys from 1999 onward, control of extreme weights through separate bounds for adjustment factors was incorporated into the GEM calibration processes for both nonresponse and poststratification. This is unlike the traditional method of winsorization in which extreme weights are truncated at prespecified levels and the trimmed portions of weights are distributed to the nontruncated cases. In GEM, it is possible to set bounds around the prespecified levels for extreme weights, and then the calibration process provides an objective way of deciding the extent of adjustment (or truncation) within the specified bounds. A step was added to poststratify the household-level weights to obtain census-consistent estimates based on the household rosters from all screened households; these household roster-based estimates then provided the control totals needed to calibrate the respondent pair weights for subsequent planned analyses. An additional step poststratified the selected person sample to conform to the adjusted roster estimates. This additional step takes advantage of the inherent two-phase nature of the NSDUH design. The final step poststratified the respondent person sample to external census data (defined within the State whenever possible, as discussed above). For more detailed information, see the section on person-level sampling weight calibration in the *2006 NSDUH Methodological Resource Book* (Chen et al., 2008).

The Reliability Study sample was embedded within the main study sample of the 2006 NSDUH. Therefore, the initial weight for the Reliability Study sample is a product of the final adjusted NSDUH main study analysis weight (ANALWT) and the inverse probability of selecting NSDUH respondents for reinterviewing in the Reliability Study. This initial weight then was adjusted through a nonresponse adjustment and a poststratification adjustment to account for bias due to nonresponse and coverage. The final analysis weight for the Reliability Study is the product of the initial weight and two adjustment factors.

One of the adjustment factors was the nonresponse adjustment. A weighting class adjustment method was used to correct weights resulting from the NSDUH respondents who were selected for the Reliability Study and who failed to complete the reinterview. Age group (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older) and race/ethnicity (Hispanic, non-Hispanic White, non-Hispanic black, and non-Hispanic others) were used to form the weighting class cells, then a cell-level adjustment factor was calculated.

The other adjustment factor was the poststratification adjustment. The Reliability Study sample excluded residents of Alaska and Hawaii, residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories), and persons who do not speak English.



Additionally, households in which two persons were selected were excluded; that is, reinterviews were conducted in households that were designated for the Reliability Study where only one person was selected. The poststratification adjustment was to reduce coverage bias and GEM was used for the poststratification. The predictor variables used in GEM were age group (12 to 17, 18 to 25, 26 to 34, 35 to 49, 50 to 64, and 65 or older), race (white, black, American Indian or Alaska Native, Asian, and two or more races), Hispanicity, gender, and some lower order interactions between these variables. The control totals for the poststratification were estimated from the census population estimates, excluding the population from Alaska and Hawaii, the population of institutional group quarters, and the population that does not speak English. To obtain the estimated control totals, a proportion of NSDUH respondents interviewed in nongroup quarters and in English was calculated using the pooled 2004, 2005, and 2006 NSDUH data for each demographic domain formed by intersecting age group, race, Hispanicity, and gender. The 2006 population estimates provided by the Census Bureau then were multiplied by these domain-specific proportions.

Estimates were computed using a multiprocedure package, SUDAAN<sup>®</sup> Software for Statistical Analysis of Correlated Data. SUDAAN was designed for the statistical analysis of data collected using stratified, multistage cluster sampling designs, as well as other observational and experimental studies involving repeated measures or studies subject to cluster correlation effects (RTI International, 2004).



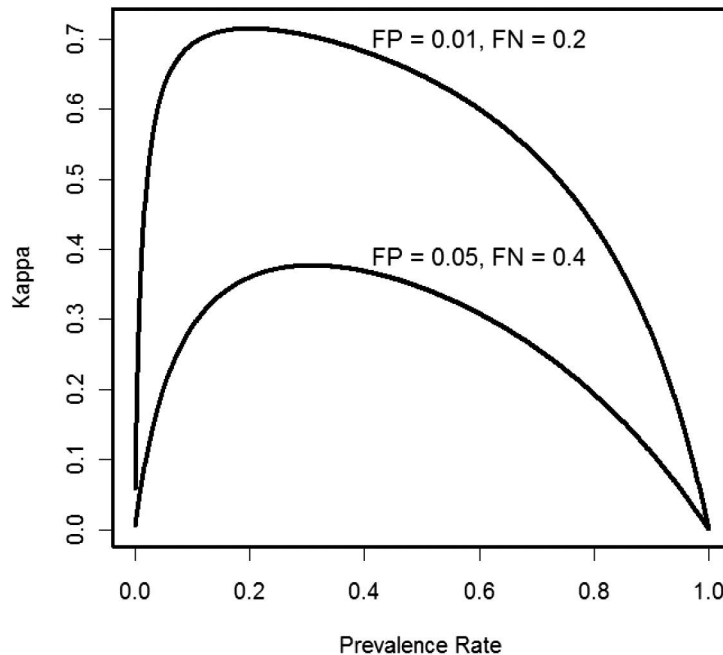
# Appendix B: Statistical Procedures

This appendix provides additional information on several statistics and related issues discussed in this report's chapters: kappa dependence on prevalence rates, computation of kappa standard errors, the index of inconsistency (*IOI*), and the kappa suppression rule.

## B.1. Kappa Dependence on Prevalence Rates

Raw agreement rates can be misleading because most of the agreement can be due to chance agreement. For example, assume that 99 percent of the responses to a "yes/no" question are "no" at either the initial interview, T1, or the follow-up interview, T2, and assume that the responses at T2 are independent of those given at T1. Then the probability of agreement under the independence assumption (referred to as "chance agreement") is  $0.99^2 + 0.01^2$ , which is approximately 98 percent, even though the agreement in this case is only due to chance. Because Cohen's kappa ( $\kappa$ ) corrects for chance agreement, it is a widely used measure (Cohen, 1960). However, Cohen's kappa is known to be sensitive to the prevalence rate, and its behavior when the prevalence rates are very close to the endpoints of the  $[0, 1]$  range should be kept in mind. For example, if the false positive (FP) rate is 1 percent and the false negative (FN) rate is 20 percent, the value of kappa can range from near 0 to about 0.7 as a function of the true prevalence (Figure B.1).

Figure B.1. Kappa Dependence on Prevalence Rates



Thompson and Walter (1988) mentioned that the raw percentage agreement "has long been recognized as a potentially misleading index" because of chance agreement, for which Cohen's kappa corrects. They discussed the issues with kappa's behavior near the endpoints 0 and

1. However, they cited Shrout, Spitzer, and Fleiss (1987) who maintained that "the dependence of kappa on prevalence may in fact be a desirable property."

## **B.2. Computation of Kappa Standard Errors**

The common approach to variance estimation for the kappa measure is to use the Fleiss, Cohen, and Everitt (1969) asymptotic variance formula (see also Agresti, 2002). It assumes an independent sample with equal probabilities of inclusion. The NSDUH Reliability Study sample design is complex, involving stratification, clustering, and unequal weighting to account for variable probabilities of inclusion and nonresponse adjustments. This may have a significant effect on the point estimates of kappa and the estimation of its variance.

Point estimates of kappa corrected for the design are straightforward. These are obtained by using weighted estimates of the proportions  $p_{00}$ ,  $p_{01}$ ,  $p_{10}$ , and  $p_{11}$  in the definition of kappa.

The variance estimate, however, is more involved. Feder (2006) developed a Taylor linearization variance estimation method for the binary case. The method subsequently was generalized to all other types of categorical variables, including ordinal variables where both exact and relaxed agreement kappas were calculated. The method was implemented in SAS<sup>®</sup> and SUDAAN<sup>®</sup> macros and used to derive all of the standard errors in this report.

The method essentially first calculates a first-order Taylor approximation to the kappa. The variance of the mean of residuals from the approximation then is estimated using SUDAAN. The variance of this mean is used as an estimate of the variance of the kappa. Feder (2006) showed by means of simulations that the approximation is good.

## **B.3. Index of Inconsistency (IOI)**

The *IOI* was used in this report to measure the inconsistencies between the initial interview (T1) and follow-up interview (T2) responses for continuous and scale variables. Low values of *IOI* indicate high reliability of responses. The advantage of using the *IOI* is that it reflects the magnitude of the differences between the T1 and T2 values. (Note that a generalized form of the kappa measure is also capable of reflecting the magnitude of differences between the measurements. However, this alternative would have required a subjective choice of weights.) The *IOI* was defined by Pritzker and Hanson (1962) as the ratio of the response variance to the total variance. The *IOI* can be written as follows:

$$IOI = \frac{\sigma_R^2}{\sigma_T^2},$$

where  $\sigma_R^2$  is half the average squared difference between potential T1 and T2 values in the population, and  $\sigma_T^2$  is the average total variance of potential responses across the population in T1 and T2. The estimation of *IOI* is given more formal expression below.

### B.3.1 Point Estimation

In this report, the *IOI* is estimated as follows: Let  $y_{i1}$  and  $y_{i2}$  be the T1 and T2 responses given by subject  $i$ . Denote by  $\bar{y}_1$  and  $\bar{y}_2$  the weighted average of  $y_{i1}$ ,  $y_{i2}$ , respectively. A nearly unbiased estimator for *IOI* is then

$$\hat{I} = \frac{\sum_{i=1}^n w_i (y_{i1} - y_{i2})^2}{\sum_{i=1}^n \sum_{t=1}^2 w_i (y_{it} - \bar{y}_t)^2}$$

(see Cochran, 1977, p. 387). In this equation,  $n$  is the sample size, and  $w_i$  is the survey weight attached to subject  $i$ .

When  $y_{i1}$  and  $y_{i2}$  have a common mean,  $y_{i+}$ , and variance,  $\sigma_{R(i)}^2$ , for all individuals in the population, the numerator of *IOI* is the average response variance in the population, hence the symbol  $\sigma_R^2$ . The denominator of *IOI* is the average variance of the  $y_{it}$  across the population (the variance of each  $y_{it}$  is the sum of  $\sigma_{R(i)}^2$  and the variance of  $y_{i+}$  around the population mean).

### B.3.2 Variance Estimation

The variance estimation of the  $\hat{I}$  in this report is based on Taylor linearization. Note that  $IOI = G / H$ , where  $G$  is the weighted sum of  $a_i = (y_{i1} - y_{i2})^2$ , and  $H$  is the weighted sum of  $b_i = (y_{i1} - \bar{y}_1)^2 + (y_{i2} - \bar{y}_2)^2$ . Therefore, the variance of  $\hat{I}$  is approximately equal to that of the total of  $r_i = (1/H)a_i - (G/H^2)b_i$ .

### B.3.3 Relationships between the *IOI* and Other Reliability Measures for Dichotomous Outcomes

Let  $y$  be a dichotomous variable with response levels 0 and 1. Let the response given by subject  $i$  at time  $t$  be  $y_{it}$ . Denote the estimated raw agreement rate by  $A$  and the estimated chance agreement by  $C$ .

Then an estimator of  $1 - \kappa$  is

$$1 - \hat{\kappa} = 1 - \frac{A - C}{1 - C} = \frac{1 - A}{1 - C} = \frac{1 - \hat{N}^{-1} \sum_{i=1}^n w_i [y_{i1} y_{i2} + (1 - y_{i1})(1 - y_{i2})]}{\bar{y}_1 + \bar{y}_2 - 2\bar{y}_1 \bar{y}_2}$$

$$= \frac{\hat{N}^{-1} \sum_{i=1}^n w_i (y_{i1} - y_{i2})^2}{\bar{y}_1 + \bar{y}_2 - 2\bar{y}_1\bar{y}_2}.$$

Or,

$$1 - \hat{\kappa} = \frac{\sum_{i=1}^n w_i (y_{i1} - y_{i2})^2}{\hat{N}(\bar{y}_1 + \bar{y}_2 - 2\bar{y}_1\bar{y}_2)}.$$

In the above formula,  $w_i$  is the survey weight,  $\bar{y}_1$  and  $\bar{y}_2$  are the weighted means at T1 and T2 (respectively), and  $\hat{N}$  is the estimated domain size.

The numerator of  $\hat{I}$  is equal to the right-hand side of the last expression for  $1 - \hat{\kappa}$  above. Thus,  $\hat{I}$  equals  $1 - \hat{\kappa}$  when  $\hat{N}(\bar{y}_1 + \bar{y}_2 - 2\bar{y}_1\bar{y}_2) = \sum_{t=1}^2 \sum_{i=1}^n w_i (y_{it} - \bar{y}_t)^2$ . Simple algebra (and using the fact that  $y_{it} = 0$  or 1 implies  $y_{it} = y_{it}^2$  for any  $i = 1, \dots, n$  and  $t = 1, 2$ ) shows that the denominators are equal if and only if  $\bar{y}_1^2 + \bar{y}_2^2 = 2\bar{y}_1\bar{y}_2$ , or, equivalently,  $\bar{y}_1 = \bar{y}_2$ . Otherwise,  $\hat{I} > 1 - \hat{\kappa}$ .

## B.4 Kappa Suppression Rule

A suppression rule was designed for this Reliability Study for the purpose of avoiding the presentation of the kappa estimate when the prevalence rate for a variable is very low (close to 0) or very high (close to 100 percent). (See Section B.1 of this appendix for a discussion of the paradoxical behavior of kappa when the prevalence is near 0 or 100.) The tables presented in Chapter 6 show a prevalence rate (weighted percent reporting condition) that is based on all responses to the initial interview, T1, whether or not the respondent completed the follow-up interview, T2. The prevalence rates used in the suppression rule require that a given respondent provide responses for both the initial and follow-up interviews. For this reason, the prevalence rate shown in the tables in Chapter 6 could be different from the prevalence rate computed for the suppression rule. Thus, there are times when the prevalence rate shown in a table is less than or equal to 1 percent or greater than or equal to 99 percent, and the associated kappa estimate is not suppressed. Similarly, the prevalence rate in the table could be greater than 1 percent or less than 99 percent and the associated kappa estimate be suppressed.

Denote the percentage of individuals responding "yes" to a dichotomous question at T1 (respectively T2) by P1 (respectively P2). In the notation of Section 4.1,  $P1 = 100 * p_{1+}$  and  $P2 = 100 * p_{+1}$ . (Note that P1 and P2 are calculated using the subsample consisting of individuals who had valid responses at both T1 and T2.) Then, the estimated kappa is suppressed when *both* conditions below are satisfied:

- $P1 \leq 1$  percent or  $P1 \geq 99$  percent or  $P2 \leq 1$  percent or  $P2 \geq 99$  percent and
- $|P1 - P2| \leq 1$  percent.

# Appendix C: Results of the T2 Interview Follow-Up Questions

For the 2006 National Survey on Drug Use and Health (NSDUH) Reliability Study, persons who completed the follow-up interview, T2, were asked an additional set of audio computer-assisted self-interviewing (ACASI) questions to gauge how much respondents felt they had remembered from the initial interview, T1, to the T2 interview, and whether their responses at the T2 interview were the same as their responses at the T1 interview. The questions and frequency of responses for all respondents and for respondents in the same versus different interviewer substudy conditions are presented in this appendix's tables. Each table shows frequencies and weighted percentages for each response level of each follow-up question by two groups defined by the number of days between T1 and T2: less than 9 days and 9 days or more. The tables also provide a *p* value for a chi-square test of the equality of the distribution across the levels between two groups: (1) those for whom there were less than 9 days between T1 and T2, and (2) those for whom T2 occurred 9 days or more after T1. A *p* value of less than or equal to .05 indicates the significance of the differences between the two groups.

Table C.1 presents the results for all respondents regardless of whether the respondent was in the same or different interviewer substudy group. Most respondents (over 77 percent) reported that they remembered most or all of their answers to the tobacco, alcohol, and marijuana use questions from the first interview. Respondents reported that most or all of their answers to the tobacco, alcohol, and marijuana use questions were the same for both interviews. There was generally greater recall of answers and greater reporting of the same answers at T1 and T2 for those persons whose T2 interview was less than 9 days after their T1 interview. However, none of the *p* values for these questions was significant.

**Table C.1. Weighted Distribution of the Responses to the Follow-Up Questions across All Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>3,136</b>	<b>100.0</b>	<b>1,904</b>	<b>55.8</b>	<b>1,232</b>	<b>44.2</b>
<b>FOLLW01 (Chi-Square <i>p</i> Value = 0.2926)</b> During the time between the first and second interviews, did you think about your use or nonuse of tobacco, alcohol, and other drugs more than usual, about the same as usual, or less than usual?						
More Than Usual	694	18.9	450	19.2	244	18.6
About the Same as Usual	2,002	66.6	1,169	64.1	833	69.8
Less Than Usual	401	13.7	262	15.8	139	11.1
Don't Know	29	0.5	15	0.6	14	0.5
Refused	8	0.2	7	0.3	1	0.0
Blank	2	0.0	1	0.0	1	0.1

(continued)

**Table C.1. Weighted Distribution of the Responses to the Follow-Up Questions across All Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study (continued)**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>3,136</b>	<b>100.0</b>	<b>1,904</b>	<b>55.8</b>	<b>1,232</b>	<b>44.2</b>
<b>FOLLW02 (Chi-Square <i>p</i> Value = 0.0569)</b> How many of the questions in this interview do you think were the <b>same</b> as the questions in the first interview?						
All of Them	1,088	27.0	684	30.5	404	22.4
Most of Them	1,774	61.6	1,068	59.5	706	64.2
Some of Them	254	11.1	143	9.6	111	12.9
None of Them	12	0.2	4	0.1	8	0.3
Don't Know	5	0.1	3	0.1	2	0.0
Refused	1	0.1	1	0.2	0	0.0
Blank	2	0.0	1	0.0	1	0.1
<b>FOLLW03 (Chi-Square <i>p</i> Value = 0.3669)</b> Please think about the questions on your use or nonuse of tobacco. How many of your answers to the tobacco questions do you remember from the <b>first</b> interview?						
All of Them	971	28.4	613	30.7	358	25.5
Most of Them	1,492	49.4	896	48.4	596	50.6
Some of Them	599	20.3	350	19.4	249	21.5
None of Them	62	1.1	38	1.3	24	0.8
Don't Know	8	0.8	4	0.2	4	1.5
Refused	2	0.0	2	0.0	0	0.0
Blank	2	0.0	1	0.0	1	0.1
<b>FOLLW04 (Chi-Square <i>p</i> Value = 0.8362)</b> [If Q3 = 1-3] How many of your answers to the tobacco questions in the second interview were the same as your answers to the tobacco questions in the first interview?						
All of Them	1,540	46.5	949	48.4	591	44.1
Most of Them	1,315	44.6	788	43.0	527	46.5
Some of Them	172	6.0	100	6.1	72	6.0
None of Them	5	0.1	3	0.1	2	0.0
Don't Know	27	0.9	17	0.8	10	1.0
Refused	3	0.0	2	0.0	1	0.0
Blank	74	1.9	45	1.6	29	2.3

(continued)



**Table C.1. Weighted Distribution of the Responses to the Follow-Up Questions across All Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study (continued)**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>3,136</b>	<b>100.0</b>	<b>1,904</b>	<b>55.8</b>	<b>1,232</b>	<b>44.2</b>
<b>FOLLW05 (Chi-Square <i>p</i> Value = 0.1806)</b> Please think about the questions on your use or nonuse of alcohol. How many of your answers to the alcohol questions do you remember from the <b>first</b> interview?						
All of Them	886	24.1	572	27.3	314	20.0
Most of Them	1,681	57.1	994	57.2	687	57.1
Some of Them	484	16.5	297	14.0	187	19.6
None of Them	68	1.4	36	1.3	32	1.4
Don't Know	12	0.9	4	0.2	8	1.7
Refused	3	0.0	0	0.0	3	0.1
Blank	2	0.0	1	0.0	1	0.1
<b>FOLLW06 (Chi-Square <i>p</i> Value = 0.5614)</b> [If Q5 = 1-3] How many of your answers to the alcohol questions in the second interview were the same as your answers to the alcohol questions in the first interview?						
All of Them	1,212	35.6	776	37.8	436	32.8
Most of Them	1,553	52.2	919	52.2	634	52.1
Some of Them	257	9.4	149	8.0	108	11.1
None of Them	12	0.1	9	0.1	3	0.1
Don't Know	16	0.4	9	0.4	7	0.5
Refused	1	0.0	1	0.0	0	0.0
Blank	85	2.3	41	1.5	44	3.3
<b>FOLLW07 (Chi-Square <i>p</i> Value = 0.7237)</b> Please think about the questions on your use or nonuse of marijuana. How many of your answers to the marijuana questions do you remember from the <b>first</b> interview?						
All of Them	1,458	50.4	905	53.2	553	46.9
Most of Them	1,248	36.2	744	35.6	504	37.0
Some of Them	345	9.7	206	9.1	139	10.5
None of Them	70	2.7	43	2.0	27	3.5
Don't Know	10	0.9	5	0.1	5	2.0
Refused	3	0.0	0	0.0	3	0.1
Blank	2	0.0	1	0.0	1	0.1

(continued)

**Table C.1. Weighted Distribution of the Responses to the Follow-Up Questions across All Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study (continued)**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>3,136</b>	<b>100.0</b>	<b>1,904</b>	<b>55.8</b>	<b>1,232</b>	<b>44.2</b>
<b>FOLLW08 (Chi-Square <i>p</i> Value = 0.3084) [If Q7 = 1-3] How many of your answers to the marijuana questions in the second interview were the same as your answers to the marijuana questions in the first interview?</b>						
All of Them	1,964	66.7	1,197	64.7	767	69.2
Most of Them	889	25.3	538	29.0	351	20.6
Some of Them	175	3.8	108	3.6	67	4.0
None of Them	10	0.5	4	0.5	6	0.5
Don't Know	11	0.1	6	0.1	5	0.1
Refused	2	0.0	2	0.0	0	0.0
Blank	85	3.7	49	2.1	36	5.6
<b>FOLLW09 (Chi-Square <i>p</i> Value = 0.1048) Now think about all of the questions in both interviews. Overall, would you say that your answers were more accurate in the first interview, more accurate in the second interview, or about as accurate each time?</b>						
More Accurate 1st Interview	230	5.2	148	6.1	82	4.0
More Accurate 2nd Interview	620	18.1	393	20.4	227	15.1
Accurate in Each Interview	2,263	75.7	1,350	72.7	913	79.4
Don't Know	19	1.1	12	0.8	7	1.4
Refused	2	0.0	0	0.0	2	0.0
Blank	2	0.0	1	0.0	1	0.1

Note: All *p* values exclude the categories "don't know," "refused," and "blank."

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

Table C.2 presents the results of the follow-up questions for those respondents who were interviewed at both the T1 and T2 interviews by the same interviewer. Similar to Table C.1, the frequencies and percentages are computed for each response level of each follow-up question by the number of days between T1 and T2.

Over 75 percent of respondents interviewed both times by the same interviewer reported that they remembered most or all of their answers to the tobacco, alcohol, and marijuana use questions from the first interview. Over 88 percent of the respondents reported that most or all of their answers to the tobacco, alcohol, and marijuana use questions were the same for both interviews. This subset of respondents exhibited the same general trend of greater recall of answers and greater reporting of the same answers at T1 and T2 when the T2 interview was less than 9 days after the T1 interview. However, the chi-square test showed that the question about remembering alcohol answers was significantly different for the less than 9 days and the 9 days or more groups, but not for tobacco or marijuana.

Also, chi-square  $p$  values for two other questions showed a significant difference between the less than 9 days group and the 9 days or more group. One question focused on how many questions the respondent thought were the same in both interviews (FOLLW02). The percentages indicate that those persons interviewed the second time less than 9 days after the first interview by the same interviewer were more confident that the questions were the same in the two interviews.

The other question that obtained significance focused on the relative accuracy of their responses for the two interviews (FOLLW09). A larger proportion of the less than 9 day respondents felt that either their first or second interview was more accurate than the other. A larger proportion of the 9 day or more respondents felt there was no difference in the accuracy of their answers to the two interviews.

**Table C.2. Weighted Distribution of Responses to the Follow-Up Questions among Same Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>2,070</b>	<b>100.0</b>	<b>1,366</b>	<b>61.4</b>	<b>704</b>	<b>38.6</b>
<b>FOLLW01 (Chi-Square <math>p</math> Value = 0.3518)</b> During the time between the first and second interviews, did you think about your use or nonuse of tobacco, alcohol, and other drugs more than usual, about the same as usual, or less than usual?						
More Than Usual	467	21.0	329	20.1	138	22.4
About the Same as Usual	1,300	63.6	826	62.1	474	66.1
Less Than Usual	281	14.7	196	16.8	85	11.3
Don't Know	14	0.5	8	0.6	6	0.2
Refused	7	0.2	6	0.4	1	0.0
Blank	1	0.0	1	0.0	0	0.0
<b>FOLLW02 (Chi-Square <math>p</math> Value = 0.0193)</b> How many of the questions in this interview do you think were the <b>same</b> as the questions in the first interview?						
All of Them	716	25.7	485	30.2	231	18.5
Most of Them	1,172	62.9	770	60.2	402	67.1
Some of Them	169	11.1	104	9.2	65	14.1
None of Them	9	0.1	4	0.1	5	0.2
Don't Know	2	0.0	1	0.0	1	0.0
Refused	1	0.2	1	0.3	0	0.0
Blank	1	0.0	1	0.0	0	0.0

(continued)

**Table C.2. Weighted Distribution of Responses to the Follow-Up Questions among Same Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study (continued)**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>2,070</b>	<b>100.0</b>	<b>1,366</b>	<b>61.4</b>	<b>704</b>	<b>38.6</b>
<b>FOLLW03 (Chi-Square <i>p</i> Value = 0.7439)</b> Please think about the questions on your use or nonuse of tobacco. How many of your answers to the tobacco questions do you remember from the <b>first</b> interview?						
All of Them	641	27.2	430	29.3	211	24.0
Most of Them	990	48.7	656	48.7	334	48.8
Some of Them	389	22.4	247	20.9	142	24.7
None of Them	43	1.0	28	1.0	15	1.0
Don't Know	5	0.7	3	0.2	2	1.5
Refused	1	0.0	1	0.0	0	0.0
Blank	1	0.0	1	0.0	0	0.0
<b>FOLLW04 (Chi-Square <i>p</i> Value = 0.5858)</b> [If Q3 = 1-3] How many of your answers to the tobacco questions in the second interview were the same as your answers to the tobacco questions in the first interview?						
All of Them	991	42.9	662	46.2	329	37.7
Most of Them	891	48.6	585	46.2	306	52.2
Some of Them	113	5.8	69	5.7	44	5.8
None of Them	4	0.1	2	0.1	2	0.1
Don't Know	19	1.0	13	0.5	6	1.7
Refused	2	0.0	2	0.1	0	0.0
Blank	50	1.7	33	1.2	17	2.5
<b>FOLLW05 (Chi-Square <i>p</i> Value = 0.0414)</b> Please think about the questions on your use or nonuse of alcohol. How many of your answers to the alcohol questions do you remember from the <b>first</b> interview?						
All of Them	575	24.4	405	28.6	170	17.6
Most of Them	1,125	55.1	724	55.6	401	54.4
Some of Them	318	18.5	210	14.9	108	24.3
None of Them	44	1.3	24	0.9	20	2.0
Don't Know	6	0.7	2	0.0	4	1.7
Refused	1	0.0	0	0.0	1	0.0
Blank	1	0.0	1	0.0	0	0.0

(continued)

**Table C.2. Weighted Distribution of Responses to the Follow-Up Questions among Same Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study (continued)**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>2,070</b>	<b>100.0</b>	<b>1,366</b>	<b>61.4</b>	<b>704</b>	<b>38.6</b>
<b>FOLLW06 (Chi-Square <i>p</i> Value = 0.4554) [If Q5 = 1-3] How many of your answers to the alcohol questions in the second interview were the same as your answers to the alcohol questions in the first interview?</b>						
All of Them	785	34.8	541	37.5	244	30.6
Most of Them	1,046	53.5	676	54.3	370	52.3
Some of Them	167	9.3	106	6.9	61	13.1
None of Them	9	0.2	7	0.1	2	0.2
Don't Know	10	0.2	8	0.3	2	0.1
Refused	1	0.0	1	0.0	0	0.0
Blank	52	2.0	27	0.9	25	3.7
<b>FOLLW07 (Chi-Square <i>p</i> Value = 0.6165) Please think about the questions on your use or nonuse of marijuana. How many of your answers to the marijuana questions do you remember from the first interview?</b>						
All of Them	945	49.0	629	51.0	316	45.8
Most of Them	836	36.8	555	38.5	281	34.2
Some of Them	234	10.1	149	9.5	85	11.1
None of Them	46	3.0	29	1.1	17	6.1
Don't Know	7	1.1	3	0.0	4	2.7
Refused	1	0.0	0	0.0	1	0.0
Blank	1	0.0	1	0.0	0	0.0
<b>FOLLW08 (Chi-Square <i>p</i> Value = 0.1189) [If Q7 = 1-3] How many of your answers to the marijuana questions in the second interview were the same as your answers to the marijuana questions in the first interview?</b>						
All of Them	1,282	65.8	844	64.1	438	68.5
Most of Them	599	25.8	401	30.5	198	18.5
Some of Them	116	3.9	78	4.2	38	3.3
None of Them	7	0.3	2	0.0	5	0.7
Don't Know	9	0.1	6	0.1	3	0.1
Refused	2	0.0	2	0.0	0	0.0
Blank	55	4.1	33	1.1	22	8.9

(continued)

**Table C.2. Weighted Distribution of Responses to the Follow-Up Questions among Same Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study (continued)**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>2,070</b>	<b>100.0</b>	<b>1,366</b>	<b>61.4</b>	<b>704</b>	<b>38.6</b>
<b>FOLLW09 (Chi-Square <i>p</i> Value = 0.0129)</b> Now think about all of the questions in both interviews. Overall, would you say that your answers were more accurate in the first interview, more accurate in the second interview, or about as accurate each time?						
More Accurate 1st Interview	177	5.6	123	6.7	54	3.8
More Accurate 2nd Interview	398	16.1	276	19.7	122	10.3
Accurate in Each Interview	1,481	77.4	957	73.1	524	84.3
Don't Know	12	0.9	9	0.5	3	1.5
Refused	1	0.0	0	0.0	1	0.0
Blank	1	0.0	1	0.0	0	0.0

Note: All *p* values exclude the categories "don't know," "refused," and "blank."

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

Table C.3 presents the results of the follow-up questions for those respondents who were interviewed by different interviewers at the T1 and T2 interviews. Similar to Tables C.1 and C.2, the frequencies and percentages are computed for each response level of each follow-up question by the number of days between T1 and T2.

More than 78 percent of the respondents interviewed by different interviewers reported that they remembered most or all of their answers to the tobacco, alcohol, and marijuana use questions from the first interview. Over 84 percent of respondents reported that most or all of their answers to the tobacco, alcohol, and marijuana use questions were the same for both interviews. This subset of respondents exhibited the same general trend of greater recall of answers and greater reporting of the same answers at T1 and T2 when the T2 interview was less than 9 days after the T1 interview. However, none of the *p* values for the chi-square tests for the questions was significant.

Overall, responses to the follow-up questions showed a trend of greater recall of answers and greater repeating of answers for those whose T2 interview was less than 9 days after the first interview than for those whose T2 interview was 9 or more days after the first interview. However, there were not enough significant chi-square tests to show statistical significance of this observed trend.

**Table C.3. Weighted Distribution of Responses to the Follow-Up Questions among Different Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>1,066</b>	<b>100.0</b>	<b>538</b>	<b>45.4</b>	<b>528</b>	<b>54.6</b>
<b>FOLLW01 (Chi-Square <i>p</i> Value = 0.4025)</b> During the time between the first and second interviews, did you think about your use or nonuse of tobacco, alcohol, and other drugs more than usual, about the same as usual, or less than usual?						
More Than Usual	227	15.8	121	18.1	106	14.0
About the Same as Usual	702	73.3	343	70.3	359	75.8
Less Than Usual	120	9.6	66	11.1	54	8.4
Don't Know	15	0.8	7	0.3	8	1.2
Refused	1	0.1	1	0.2	0	0.0
Blank	1	0.3	0	0.0	1	0.6
<b>FOLLW02 (Chi-Square <i>p</i> Value = 0.5010)</b> How many of the questions in this interview do you think were the <b>same</b> as the questions in the first interview?						
All of Them	372	31.0	199	31.5	173	30.6
Most of Them	602	56.8	298	58.0	304	55.8
Some of Them	85	11.5	39	10.3	46	12.5
None of Them	3	0.2	0	0.0	3	0.5
Don't Know	3	0.1	2	0.2	1	0.0
Refused	0	0.0	0	0.0	0	0.0
Blank	1	0.3	0	0.0	1	0.6
<b>FOLLW03 (Chi-Square <i>p</i> Value = 0.0814)</b> Please think about the questions on your use or nonuse of tobacco. How many of your answers to the tobacco questions do you remember from the <b>first</b> interview?						
All of Them	330	28.8	183	33.6	147	24.7
Most of Them	502	49.7	240	46.3	262	52.4
Some of Them	210	18.0	103	15.0	107	20.6
None of Them	19	2.7	10	4.9	9	1.0
Don't Know	3	0.5	1	0.2	2	0.7
Refused	1	0.0	1	0.0	0	0.0
Blank	1	0.3	0	0.0	1	0.6

(continued)

**Table C.3. Weighted Distribution of Responses to the Follow-Up Questions among Different Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study (continued)**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>1,066</b>	<b>100.0</b>	<b>538</b>	<b>45.4</b>	<b>528</b>	<b>54.6</b>
<b>FOLLW04 (Chi-Square <i>p</i> Value = 0.2523)</b> [If Q3 = 1-3] How many of your answers to the tobacco questions in the second interview were the same as your answers to the tobacco questions in the first interview?						
All of Them	549	49.9	287	54.1	262	46.4
Most of Them	424	37.9	203	33.1	221	41.9
Some of Them	59	7.6	31	5.9	28	9.0
None of Them	1	0.1	1	0.2	0	0.0
Don't Know	8	1.0	4	1.7	4	0.4
Refused	1	0.0	0	0.0	1	0.0
Blank	24	3.5	12	5.1	12	2.3
<b>FOLLW05 (Chi-Square <i>p</i> Value = 0.6960)</b> Please think about the questions on your use or nonuse of alcohol. How many of your answers to the alcohol questions do you remember from the <b>first</b> interview?						
All of Them	311	23.1	167	24.8	144	21.7
Most of Them	556	56.6	270	53.8	286	59.0
Some of Them	166	15.6	87	15.9	79	15.3
None of Them	24	3.5	12	5.1	12	2.2
Don't Know	6	0.8	2	0.5	4	1.0
Refused	2	0.1	0	0.0	2	0.2
Blank	1	0.3	0	0.0	1	0.6
<b>FOLLW06 (Chi-Square <i>p</i> Value = 0.2091)</b> [If Q5 = 1-3] How many of your answers to the alcohol questions in the second interview were the same as your answers to the alcohol questions in the first interview?						
All of Them	427	34.2	235	40.2	192	29.2
Most of Them	507	50.6	243	44.7	264	55.5
Some of Them	90	9.6	43	8.8	47	10.2
None of Them	3	0.0	2	0.0	1	0.0
Don't Know	6	0.9	1	0.7	5	1.1
Refused	0	0.0	0	0.0	0	0.0
Blank	33	4.7	14	5.6	19	4.0

(continued)



**Table C.3. Weighted Distribution of Responses to the Follow-Up Questions among Different Interviewers, by Number of Days between T1 and T2: 2006 NSDUH Reliability Study (continued)**

Question	Total		Less Than 9 Days		9 Days or More	
	#	%	#	%	#	%
<b>Total</b>	<b>1,066</b>	<b>100.0</b>	<b>538</b>	<b>45.4</b>	<b>528</b>	<b>54.6</b>
<b>FOLLW07 (Chi-Square <i>p</i> Value = 0.0651)</b> Please think about the questions on your use or nonuse of marijuana. How many of your answers to the marijuana questions do you remember from the <b>first</b> interview?						
All of Them	513	51.1	276	54.8	237	48.0
Most of Them	412	34.3	189	29.2	223	38.6
Some of Them	111	10.5	57	9.3	54	11.6
None of Them	24	3.3	14	6.5	10	0.6
Don't Know	3	0.3	2	0.2	1	0.5
Refused	2	0.1	0	0.0	2	0.2
Blank	1	0.3	0	0.0	1	0.6
<b>FOLLW08 (Chi-Square <i>p</i> Value = 0.1677)</b> [If Q7 = 1-3] How many of your answers to the marijuana questions in the second interview were the same as your answers to the marijuana questions in the first interview?						
All of Them	682	66.2	353	67.1	329	65.5
Most of Them	290	24.2	137	23.4	153	24.8
Some of Them	59	5.1	30	2.0	29	7.8
None of Them	3	0.4	2	0.8	1	0.1
Don't Know	2	0.0	0	0.0	2	0.0
Refused	0	0.0	0	0.0	0	0.0
Blank	30	4.1	16	6.8	14	1.8
<b>FOLLW09 (Chi-Square <i>p</i> Value = 0.1248)</b> Now think about all of the questions in both interviews. Overall, would you say that your answers were more accurate in the first interview, more accurate in the second interview, or about as accurate each time?						
More Accurate 1st Interview	53	4.2	25	2.7	28	5.3
More Accurate 2nd Interview	222	22.2	117	26.1	105	19.0
Accurate in Each Interview	782	72.0	393	69.7	389	73.8
Don't Know	7	1.3	3	1.5	4	1.1
Refused	1	0.0	0	0.0	1	0.0
Blank	1	0.3	0	0.0	1	0.6

Note: All *p* values exclude the categories "don't know," "refused," and "blank."

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2006 Reliability Study (*n* = 3,136).

In general, responses to the follow-up questions were similar for both the same interviewer substudy respondents and different interviewer substudy respondents. The total percentages typically varied up to three or four percentage points between the two groups. The largest difference observed was how often respondents reported thinking about their use or nonuse of tobacco, alcohol, and other drugs between the first and second interviews. Approximately 64 percent of respondents interviewed by the same interviewer thought about their use or nonuse about the same as usual compared with 73 percent of respondents interviewed by a different interviewer. In contrast, more same interviewer substudy respondents (21 percent) than different interviewer substudy respondents (16 percent) thought about their use or nonuse more than usual between the first and second interviews.

Another interesting observation is for the last question that asks whether respondents felt their answers were more accurate during the first or second interview or whether their answers had about the same accuracy during both interviews. Of those respondents who had the same interviewer both times, 77 percent felt their answers were accurate during both interviews and 16 percent felt their answers during the second interview were more accurate. For those respondents who had a different interviewer the second time, only 72 percent felt the accuracy of their responses was the same during both interviews, and 22 percent felt their responses were more accurate during the second interview.

# Appendix D: List of Contributors

This National Survey on Drug Use and Health (NSDUH) report was prepared by the Division of Population Surveys, Office of Applied Studies (OAS), Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services (HHS), and by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina. Work by RTI was performed under Contract No. 283-2004-00022.

At SAMHSA, the following staff in alphabetical order were coauthors of the report: Joseph C. Gfroerer and Joel Kennet. At RTI, the following staff in alphabetical order were responsible for writing the report: James R. Chromy, Moshe Feder, Erica Hirsch, Katherine B. Morton, Lanny Piper, Beth H. Riggsbee, Jeanne A. Snodgrass, Thomas G. Virag, and Feng Yu.

Also at SAMHSA, Joel Kennet was responsible for oversight of planning and production of the report, and Joseph C. Gfroerer, Arthur Hughes, and Michael Jones reviewed and provided editorial suggestions. Also at RTI, the following staff were responsible for generating the estimates and providing other support and analysis: Erica Hirsch, Feng Yu, Lisa E. Packer, Paul P. Biemer, and Brenda Porter. Lanny Piper was task leader for production of the report. Phillip S. Kott and Mary Ellen Marsden provided review comments, and Mary Ellen Marsden provided oversight for final report production. Richard Straw edited the report, Joyce Clay-Brooks formatted its text and tables, and Teresa Bass, Cassandra Carter, Kim Cone, E. Andrew Jessup, Shari B. Lambert, Pamela Couch Prevatt, and Cheryl Velez prepared its print and Web versions. Diane Philyaw provided assistance with production of the figures. Final report production was provided by Christine Hagar, Jane Feldman, and Coleen Sanderson at SAMHSA.

