

ICPSR 34861

## **Monitoring the Future: A Continuing Study of American Youth (12th-Grade Survey), 2012**

Lloyd D. Johnston

*University of Michigan. Institute for Social  
Research. Survey Research Center*

Jerald G. Bachman

*University of Michigan. Institute for Social  
Research. Survey Research Center*

Patrick M. O'Malley

*University of Michigan. Institute for Social  
Research. Survey Research Center*

John E. Schulenberg

*University of Michigan. Institute for Social  
Research. Survey Research Center*

Codebook for Form 4 Data

Inter-university Consortium for  
Political and Social Research  
P.O. Box 1248  
Ann Arbor, Michigan 48106  
[www.icpsr.umich.edu](http://www.icpsr.umich.edu)

# Terms of Use

The terms of use for this study can be found at:  
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/34861/terms>

## Information about Copyrighted Content

Some instruments administered as part of this study may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

### NOTICE

#### WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

## **INTRODUCTION**

### **DATA COLLECTION DESCRIPTION**

MONITORING THE FUTURE: A CONTINUING STUDY OF AMERICAN YOUTH, 2012 is conducted by the University of Michigan's Institute for Social Research and receives its core funding under grants from the National Institute on Drug Abuse. (The responsible investigators are: Lloyd D. Johnston, principal investigator; Jerald G. Bachman, Patrick M. O'Malley, and John Schulenberg, co-principal investigators.) The research project is unusually comprehensive in several respects: surveys are conducted annually on an ongoing basis; the samples are large and nationally representative; and the subject matter is very broad, encompassing some 1400 variables per year.

The Monitoring the Future Project is designed to explore changes in many important values, behaviors, and lifestyle orientations of contemporary American youth. Two general types of tasks may be distinguished. The first is to provide a systematic and accurate "description" of the youth population of interest in a given year, and to quantify the direction and rate of the changes taking place among them over time. The second task, more analytic than descriptive, involves the "explanation" of the relationships and trends observed to exist.

### **DATA COLLECTION PROCEDURES**

The basic research design involves annual data collections from high school seniors during the spring of each year, beginning with the class of 1975. Each data collection takes place in approximately 130 public and private high schools selected to provide an accurate cross-section of high school seniors throughout the United States.

One limitation in the design is that it does not include in the target population those young men and women who drop out of high school before graduation (or before the last few months of the senior year, to be more precise). This excludes a relatively small proportion of each age cohort -- between 11 and 20 percent -- though not an unimportant segment, since certain behaviors, such as illicit drug use and delinquency tend to be higher than average in this group. However, the addition of a representative sample of dropouts would increase the cost of the present research enormously, because of their dispersion and generally higher level of resistance to being located and interviewed.

For the purposes of estimating characteristics of the entire age group, the omission of high school dropouts does introduce certain biases; however, their small proportion sets outer limits on the bias. For the purposes of estimating "changes" from one cohort of high school seniors to another, the omission of dropouts represents a problem only if different cohorts have considerably different proportions who drop out. There is no reason to expect dramatic changes in those rates for the foreseeable future, and recently published government statistics indicate only very small decreases in dropout rates since 1970.

Some may use this high school data to draw conclusions about changes for the entire age group. While the investigators do not encourage such extrapolation, they suspect that the conclusions reached often would be valid, since over 80 percent of the age group is in the surveyed segment of the population and changes among those not in school are likely to parallel the changes among those who are.

### **SAMPLING INFORMATION**

The procedure for securing a nationally representative sample of high school seniors in public and private schools is a multi-stage one. Stage 1 is the selection of particular geographic areas, Stage 2 is the selection of one or more high schools in each area, and Stage 3 is the selection of seniors within each high school.

**STAGE 1: GEOGRAPHIC AREAS.** The geographic areas used in this study are the primary sampling units (PSUs) developed by the Sampling Section of the Survey Research Center for use in the Center's nationwide interview studies. Because these same PSUs are used for personal interview studies by the Survey Research Center (SRC), local field representatives can be assigned to administer the data collections in practically all schools.

**STAGE 2: SCHOOLS.** In the major metropolitan areas more than one high school is often included in the sampling design; in most other sampling areas a single high school is sampled. In all cases, the selections of high schools are made such that the probability of drawing a school is proportionate to the size of its senior class. When a sampled school is unwilling to participate, a replacement school as similar to it as possible is selected from the same geographic area.

**STAGE 3: STUDENTS.** Within each selected school, up to about 350 seniors may be included in the data collection. In schools with fewer than 350 seniors, we attempt to include all of them in the data collection. In larger schools, a subset of seniors is selected either by randomly sampling classrooms or by some other random method that is convenient for the school and judged to be unbiased. A sampling weight is assigned to each respondent so as to take account of variations in the sizes of samples from one school to another, as well as the variations in selection probabilities occurring at the earlier stages of sampling. For a table of the sample size and student response rates see Appendix B.

One other important feature of the base-year sampling procedure should be noted here. All schools (except for half of the initial 1975 sample) are asked to participate in two data collections, thereby permitting replacement of half of the total sample of schools each year. One motivation for requesting that schools participate for two years is administrative efficiency; it is a costly and time-consuming procedure to secure the cooperation of schools, and a two-year period of participation cuts down that effort substantially. Another important advantage is that whenever an appreciable shift in scores from one graduating class to the next is observed, it is possible to check whether the shift might be attributable to some differences in the newly sampled schools. This is done simply by repeating the analysis using only the 60 or so schools which participated both years. Thus far, the half-sample approach has worked quite well and

examination of drug prevalence data from the "matched half-samples" showed that the half samples of repeat schools yielded drug prevalence trends which were virtually identical to trends based on all schools.

**SCHOOL RECRUITING PROCEDURES.** Early during the fall semester an initial contact is made with each sampled school. First, a letter is sent to the principal describing the study and requesting permission to survey seniors. The letter is followed by a telephone call from a project staff member, who attempts to deal with any questions or problems and (when necessary) makes arrangements to contact and seek permission from other school district officials. Basically the same procedures are followed for schools asked to participate for the second year.

Once the school's agreement to participate is obtained, arrangements are made by phone for administering the questionnaires. A local SRC representative is assigned to work with the school to set a mutually agreeable date for the survey and to carry out the administration.

**ADVANCE CONTACT WITH TEACHERS AND STUDENTS.** The local SRC representative is instructed to visit the school two weeks ahead of the actual date of administration. This visit serves as an occasion to meet the teachers whose classes will be affected and to provide them with a brochure describing the study, a brief set of guidelines about the questionnaire administration, and a supply of flyers to be distributed to the students a week to 10 days in advance of the questionnaire administration. The guidelines to the teachers include a suggested announcement to students at the time the flyers are distributed.

From the students' standpoint, the first information about the study usually consists of the teacher's announcement and the short descriptive flyer. In announcing the study, the teachers are asked to stress that the questionnaires used in the survey are not tests, and that there are no right or wrong answers. The flyer tells the students that they will be invited to participate in the study, points out that their participation is strictly voluntary, and stresses confidentiality (including a reference to the fact that the Monitoring the Future project has a special government grant of confidentiality which allows their answers to be protected). The flyer also serves as an informative document which the students can show to their parents. Parental consent involves, at a minimum, the school mailing a letter describing the study and a copy of the student flyer to the parents. The letter provides parents with an easy way to decline their child's participation, if they so wish. Active consent procedures are used when the school or district requires them.

**QUESTIONNAIRE ADMINISTRATION.** The questionnaire administration in each school is carried out by the local SRC representatives and their assistants, following standardized procedures detailed in a project instruction manual. The questionnaires are administered in classrooms during normal class periods whenever possible, although circumstances in some schools require the use of larger group administrations. Teachers are not asked to do anything more than introduce the SRC staff members and (in most cases) remain in the classroom to help guarantee an orderly atmosphere for the survey. Teachers are urged to avoid walking around the room, so that students may feel free to write their answers without fear of being observed.

The actual process of completing the questionnaires is quite straightforward.

Respondents are given sharpened pencils and asked to use them because the questionnaires are designed for automated scanning. Most respondents can finish within a 45 minute class period; for those who cannot, an effort is made to provide a few minutes of additional time.

**PROCEDURES FOR PROTECTING CONFIDENTIALITY.** In any study that relies on voluntary reporting of drug use or other illegal acts, it is essential to develop procedures which guarantee the confidentiality of such reports. It is also desirable that these procedures be described adequately to respondents so that they are comfortable about providing honest answers.

The first information given to students about the survey consists of a descriptive flyer stressing the confidentiality and voluntary participation. This theme is repeated at the start of the questionnaire administration. Each participating student is instructed to read the message on the cover of the questionnaire, which stresses the importance and value of the study, notes that answers will be kept strictly confidential, states that the study is completely voluntary, and tells the student "If there is any question you or your parents would find objectionable for any reason, just leave it blank." The instructions then point out that in a few months a summary of nationwide results will be mailed to all participants and also that a follow-up questionnaire will be sent to some students after a year. The cover message explains that these are the reasons for asking that name and address be written on a special form which will be removed from the questionnaire and handed in separately. The message also points out that the two different code numbers (one on the questionnaire and one on the tear-out form) cannot be matched except by a special computer file at the University of Michigan.

In order to protect the confidentiality of responses and the identity of respondents, a number of alterations have been made in the original dataset to prepare it for public release; these alterations are described later in the section "Processing Information."

## **CONTENT AREAS AND QUESTIONNAIRE DESIGN**

Drug use and related attitudes are the topics which receive the most extensive coverage in the Monitoring the Future project; but the questionnaires also deal with a wide range of other subject areas, including attitudes about government, social institutions, race relations, changing roles for women, educational aspirations, occupational aims, and marital and family plans, as well as a variety of background and demographic factors.

The following table shows the subject area codes and definitions which are used in the [cross-time index](#) of base year grade 12 questionnaire items provided separately in this archive.

### **MEASUREMENT CONTENT AREAS**

---

---

- A. **DRUGS.** Drug use and related attitudes and beliefs, drug availability and exposure, surrounding conditions and social meaning of drug use. Views of significant others

regarding drugs.

- B. EDUCATION. Educational lifestyle, values, experiences, and environments
- C. WORK AND LEISURE. Vocational values, meaning of work and leisure, work and leisure activities including computer use, preferences regarding occupational characteristics and type of work setting.
- D. SEX ROLES AND FAMILY. Values, attitudes, and expectations about marriage, family structure, sex roles, and sex discrimination.
- E. POPULATION CONCERNS. Values and attitudes about overpopulation and birth control.
- F. CONSERVATION, MATERIALISM, EQUITY, ETC. Values, attitudes, and expectations related to conservation, pollution, materialism, equity, and the sharing of resources. Preferences regarding type of dwelling and urbanicity.
- G. RELIGION. Religious affiliation, practices, and views.
- H. POLITICS. Political affiliation, activities, and views.
- I. SOCIAL CHANGE. Values, attitudes, and expectations about social change.
- J. SOCIAL PROBLEMS. Concern with various social problems facing the nation and the world.
- K. MAJOR SOCIAL INSTITUTIONS. Confidence in and commitment to various major social institutions (business, unions, branches of government, press, organized religion, military, etc.).
- L. MILITARY. Views about the armed services and the use of military force. Personal plans for military service.
- M. INTERPERSONAL RELATIONSHIPS. Qualitative and quantitative characteristics of cross-age and peer relationships. Interpersonal conflict.
- N. RACE RELATIONS. Attitudes toward and experiences with other racial groups.
- O. CONCERN FOR OTHERS. Concern for others; voluntary and charitable activities.
- P. HAPPINESS. Happiness and life satisfaction, overall and in specific life domains.
- Q OTHER PERSONALITY VARIABLES. Attitudes about self (including self-esteem), locus of control, loneliness, risk-taking, trust in others, importance placed on various life goals, counterculture orientation, hostility, boredom.
- R. BACKGROUND. Demographic and family background characteristics, living arrangements.
- S. DEVIANT BEHAVIOR AND VICTIMIZATION. Delinquent behaviors, driving violations and accidents (including those under the influence of drugs), victimization experiences.
- T. HEALTH. Health habits, somatic symptoms, illness, medical treatment.

---

---

Given this breadth of content, the study is not presented to respondents as a "drug use study," nor do they tend to view it as such.

Because many questions are needed to cover all of these topic areas, much of the questionnaire content is divided into different questionnaire forms which are distributed to participants in an ordered sequence. (Five forms were used in 1975-88; a sixth form was added in 1989.) This sequence produces five or six virtually identical subsamples.

About one-third of each questionnaire form consists of key or "core" variables which are common to all forms. All demographic variables and some measures of drug use are included in this "core" set of measures. This use of the full sample for drug and demographic measures provides a more accurate estimation on these dimensions and also makes it possible to link them statistically to all the other measures which are included in a single form only.

### **REPRESENTATIVENESS AND VALIDITY**

The samples for this study are intended to be representative of high school seniors attending private or public schools throughout the 48 contiguous states. We have already discussed the fact that this definition of the sample excludes one important portion of the age cohort: those who have dropped out of high school before nearing the end of the senior year. But given the aim of representing high school seniors, it will now be useful to consider the extent to which the obtained samples of schools and students are likely to be representative of all seniors and the degree to which the data obtained are likely to be valid.

It is possible to distinguish at least four ways in which survey data of this sort might fall short of being fully representative. First, some sampled schools refuse to participate, which could introduce some bias. Second, the failure to obtain questionnaire data from 100 percent of the students sampled in participating schools would also introduce bias. Third, the answers provided by participating students are open to both conscious and unconscious distortions which could reduce validity. Finally, limitations in sample size and/or design could place limits on the accuracy of estimates.

**SCHOOL PARTICIPATION.** As noted in the description of the sampling design, schools are invited to participate in the study for a two-year period. For each school that declines to participate - an occurrence which happens, on average, a little over one-third of the time -- a similar school (in terms of size, geographic area, urbanicity, etc.) is recruited as a replacement for that "slot". Since the study's inception, either an original school or a replacement school has been obtained for between 95% to 99% of the sample units, or "slots". With very few exceptions, each school which has participated for one data collection has agreed to participate for a second. The selection of replacement schools almost entirely removes problems of bias in region, urbanicity, and the like that might result from certain schools refusing to participate. Other potential biases are more subtle, however. For example, if it turned out that most schools with "drug problems" refused to participate, that would seriously bias the drug estimates derived from the sample. And if any other single factor were dominant in most refusals, that also might suggest a source of serious bias. In fact, however, the reasons for schools' refusals to participate



are varied and largely a function of happenstance events of the particular year. Thus, the investigators feel fairly confident that school refusals have not seriously biased the surveys.

**STUDENT PARTICIPATION.** Completed questionnaires are obtained from three-fourths to four-fifths of all 12<sup>th</sup> graders sampled. The single most important reason that students are missed is that they are absent from class at the time of data collection, and in most cases it is not workable to schedule a special follow-up data collection for them. Students with fairly high rates of absenteeism also report above-average rates of drug use; therefore, there is some degree of bias introduced by missing the absentees. That bias could be corrected through the use of special weighting; however, this course was not chosen because the bias in estimates (in drug use, where the potential effect was hypothesized to be largest) was determined to be quite small and because the necessary weighting procedures would have introduced undesirable complications. In addition to absenteeism, student nonparticipation occurs because of schedule conflicts with school trips and other activities which tend to be more frequent than usual during the final months of the senior year. Of course, some students refuse to complete or turn in a questionnaire, either on their own or because their parents refused consent. However, SRC representatives in the field estimate this proportion to be only about two percent.

**VALIDITY OF SELF-REPORT DATA.** Survey measures of delinquency and of drug use depend upon respondents reporting what are, in many cases, illegal acts. Thus, a critical question is whether such self-reports are likely to be valid. Like most studies dealing with these areas, the present study does not include direct, objective validation of the present measures; however, the considerable amount of inferential evidence which exists strongly suggest that the self-report questions produce largely valid data. A number of factors have given the investigators reasonable confidence about the validity of the responses to what are presumably among the most sensitive questions in the study: a low non-response rate on the drug questions; a large proportion admitting to some illicit drug use; the consistency of findings across several years of the present study; strong evidence of construct validity (based on relationships observed between variables); a close match between these data and the findings from other studies using other methods; and the findings from several methodological studies which have used objective validation methods.

As for others of the measures, a few have a long and venerable history -- as scholars of the relevant literature will recognize -- though some of these measures have been modified to fit the present questionnaire format. Many questions, however, have been developed specifically for this project through a process of question writing, pilot testing, pretesting, and question revision or elimination. Some have already been included in other publications from the study, but many have not; therefore, there exists little empirical evidence of their validity and reliability.

**ACCURACY OF THE SAMPLE.** A sample survey never can provide the same level of accuracy as would be obtained if the entire target population were to participate in the survey -- in the case of the present study, about 3 – 4 million seniors per year. But perfect accuracy of this sort would be extremely expensive and certainly not worthwhile considering the fact that a high level of accuracy can be provided by a carefully designed probability sample. The accuracy of the sample in this study is affected both by the size of the student sample and by the number of

schools in which they were clustered. For the purposes of this introduction, it is sufficient to note that virtually all estimates based on the total sample have confidence intervals of +/- 1.5 percentage points or smaller - sometimes considerably smaller. This means that, had the project been able to invite all schools and all seniors in the 48 contiguous states to participate, the results from such a massive survey would be within an estimated 1.5 percentage points from the present sample findings 95 times out of 100. This is a quite high level of accuracy, and one that permits the detection of fairly small trends from one year to the next.

Because of the complex sampling design, standard means of assessing confidence intervals are not appropriate. The [annual volumes](#) from the project can provide information which allow the analyst to determine the confidence intervals around means and percentages for both the total sample and various subgroups. They also provide tables and guidelines for testing the statistical significance of differences between subgroups, and the significance of year-to-year changes.

**CONSISTENCY AND THE MEASUREMENT OF TRENDS.** One other point is worth noting in a discussion of the validity of the findings. The Monitoring the Future project is, by intention, a study designed to be sensitive to changes from one time to another. Accordingly, the measures and procedures have been standardized and applied consistently across each data collection. To the extent that any biases remain because of limits in school and/or student participation, and to the extent that there are distortions (lack of validity) in the responses of some students, it seems very likely that such problems will exist in much the same way from one year to the next. In other words, biases in the survey estimates should tend to be consistent from one year to another, which means that the measurement of trends should be affected very little by such biases.

**INTERPRETING RACIAL DIFFERENCES.** Until 2005, ethnic identification was provided for the two largest racial/ethnic subgroups in the population -- those who identified themselves as white or Caucasian and those who identified themselves as black or African American. Identification was not given for the other ethnic categories (Native Americans, Asian Americans, Mexican Americans, Puerto Rican Americans, or other Latin Americans) since each of these groups comprised a small proportion of the sample in any given year, which means that their small Ns (in combination with their clustered groupings in a limited number of schools) would yield estimates which would be too unreliable. Because of increases in the number of those who identify themselves as one of the Hispanic groups, we now include identification for this category.

However, the analyst should bear in mind that African Americans and Hispanics -- each of which constitutes approximately 8-15 percent of each year's sample -- are represented by perhaps as few as 200 respondents per year on any single questionnaire form. Further, because our sample is a stratified clustered sample, it yields less accuracy than would be yielded by a pure random sample of equal size (see Appendix B of the [annual volumes](#) for details). Therefore, because of the limited number of cases, the margin of sampling error around any statistic describing African Americans or Hispanics is larger than for most other subgroups.

There exists, however, a way to determine the replicability of any finding involving racial

comparisons. Since most questions are repeated from year to year, one can readily establish the degree to which a finding is replicated by looking at the results in prior and subsequent years. Given the relatively small Ns for minority groups, the analyst is urged to seek such replication before putting much faith in the reliability of any particular racial comparison.

There are factors in addition to reliability, however, which could be misleading in the interpretation of racial differences. Given the social importance which has been placed on various racial differences reported in the social science literature, the investigators would like to caution the analyst to consider the various factors which could account for differences. These factors fall into three categories: differential representation in the sample, differential response tendencies, and the confounding of race with a number of other background and demographic characteristics. The following discussion is based on analyses that were conducted prior to 2005, when identifiers for Hispanics were not included, so the discussion is specific to African Americans. However, the points made, particularly those about differential representation and confounding of race/ethnicity with other background and demographic characteristics, would be relevant to Hispanics, as well.

**DIFFERENTIAL REPRESENTATION.** Census data characterizing American young people in the approximate age range of those in this sample show somewhat lower proportions of African Americans than whites remain in school through the end of the twelfth grade. Therefore, a slightly different segment of the African American population than of the white population resides in the target population of high school seniors. Further, the samples appear to under represent slightly those African American males who, according to census figures, are in high school at the twelfth grade level. Identified African American males comprise about 6 percent of the sample, whereas census data suggest that they should comprise around 7 percent. Therefore it appears that more African American males are lost from the target population than white males or females of either race. This may be due to generally poorer attendance rates on the part of some African American males and/or an unwillingness on the part of some to participate in data collections of this sort.

In sum, a smaller segment of the African American population than of the white population of high school age is represented by the data contained here. Insofar as any characteristic is associated with being a school dropout or absentee, it is likely to be somewhat disproportionately underrepresented among African Americans in the sample.

**DIFFERENTIAL RESPONSE TENDENCIES.** In examining the full range of variables, racial differences in response tendencies have been noted. First, the tendency to state agreement in response to agree-disagree questions is generally somewhat greater among African Americans than among whites. For example, African Americans tend to agree more with the positively worded items in the index of self-esteem, but they also tend to agree more with the negatively worded items. As it happens, that particular index has an equal number of positively and negatively worded items, so that any overall "agreement bias" should be self-canceling when the index score is computed. However, group differences in agreement bias are likely to affect results on questions employing the agree-disagree format. Fortunately, most of the questions are not of that type.

There has also been observed a somewhat greater than average tendency for African American respondents to select extreme answer categories on attitudinal scales. For example, even if the same proportion of African Americans as whites felt positively (or negatively) about some subject, fewer of the whites are likely to say they feel very positively (or negatively). The analyst should be aware that differences in responses to particular questions may be related to these more general tendencies.

A somewhat separate issue in response tendency is a respondent's willingness to answer particular questions. The missing data rate may reflect willingness to answer particular questions. If a particular question or set of questions has a missing data rate higher than is true for the prior or subsequent questions, then presumably more respondents than usual were unwilling (or perhaps unable) to answer it. Such an exaggerated missing data rate exists for African American males on the set of questions dealing with the respondent's own use of illicit drugs. Clearly a respondent's willingness to be candid on such questions depends on his or her trust of the research process and of the researchers themselves. The exaggerated missing data rates for African American males in these sections may reflect, at least in part, less trust. The analyst is advised to check for exceptional levels of missing data when making comparisons on any variable in which candor is likely to be reduced by lower system trust. One bit of additional evidence related to trust in the research process is that higher proportions of African Americans than whites reported that if they had used marijuana or heroin they would not have been willing to report it in the survey.

**COVARIANCE WITH OTHER FACTORS.** Some characteristics such as race are highly confounded (correlated) with other variables -- variables which may in fact explain some observed racial differences. Put another way, at the aggregate level we might observe a considerable racial difference on some characteristic, but once we control for some background characteristic such as socio-economic level or region of the country -- that is, once we compare the African American respondents with whites who come from similar backgrounds -- there may be no racial difference at all.

Race is correlated with important background and demographic variables. A higher proportion of African Americans live in the South and a higher proportion grew up in families with the mother and/or father absent, and more had mothers who worked while they were growing up. A substantially higher proportion of African Americans are Baptists, and African Americans tend to attribute more importance to religion than do whites. A higher proportion of African American respondents have children, and on the average they are slightly older than the white sample. As was mentioned earlier African American males are more underrepresented in our sample than African American females.

These differences in background, demographic, and descriptive characteristics are noted because, in any attempt to understand why a racial difference exists, one would want to be able to examine the role of these covarying characteristics.

## WEIGHTING INFORMATION

Frequency and percentage distributions displayed in codebooks produced after 2007 are unweighted, rather than weighted by variable ARCHIVE\_WT (previously V5) as they had been in previous years. This change was made to simplify both the production of the codebooks and their interpretation by the analyst.

## FILE STRUCTURE

MONITORING THE FUTURE: A CONTINUING STUDY OF AMERICAN YOUTH, 2012 is available from ICPSR as seven logical record length datasets. Each dataset consists of SAS, SPSS, and Stata setup files containing all technical information for each variable in the corresponding datafile, and the datafile itself. The data are sorted by case. The datasets are organized by the form number (questionnaire version) used. For each part, the data are also available from ICPSR in the following formats: SAS transport (CPORT) file, SPSS system file, and Stata system file, with SAS and Stata supplemental syntax files, and a tab-delimited ascii text file.

part #	form	# of variables	Logical record length	Unweighted n
1	Core	134	278	14,343
2	Form 1	646	1,303	2,379
3	Form 2	331	671	2,390
4	Form 3	361	732	2,385
5	Form 4	271	553	2,396
6	Form 5	309	627	2,401
7	Form 6	339	688	2,392

The SAS, SPSS, and Stata setup files give the format and other information for each variable in the data file. See the section "Codebook Information" for further details. The data file is constructed with a single logical record for each case.

## ICPSR PROCESSING INFORMATION

The data collection was processed according to the standard ICPSR processing procedures. The data were checked for illegal or inconsistent code values which, when found, were recoded to missing data values. Consistency checks were performed.

NOTE: THE "cases" IN THE CODEBOOK INCLUDES MISSING DATA ON THE QUESTION INVOLVED.

For reasons of confidentiality, the weight variable (ARCHIVE\_WT) was altered from its original version to a modified version prior to public distribution of the data. THIS RESULTS IN SLIGHT DISCREPANCIES BETWEEN THE PERCENTAGES AND N SIZES IN THE ANNUAL ISR VOLUMES AND THOSE FROM WEIGHTED ANALYSES OF THE PUBLIC USE DATASETS. Typically, the variation is less than 1%.

In order to protect the confidentiality of responses and the identity of respondents, a number of alterations and omissions have been made in the original dataset to prepare it for public release. Three variables have been included to describe the respondent's general environment without identifying school or state. These are (1) region (Northeast, North Central, South, and West), (2) whether or not the school is located in a Metropolitan Statistical Area (MSA), and (3) whether or not the school is located in a Large MSA. Some questions have been eliminated altogether; others are collapsed to mask groups which are very small. The following table lists the question numbers and names of the variables which have been excluded from each twelfth grade dataset.

### OMITTED VARIABLES:

All datasets	C01. R'S BIRTH YEAR C02. R'S BIRTH MONTH C04A-I, R'S RACE (9 categories) C07A-B. # OLDER BR/SR, # YOUNGER BR/SR C07Ca,e-i. R'S HSHLD (other than mother/father/sibling) C13A. R'S RELGS PRFNC
Form 1	D19. CURRENT HEIGHT D20. CURRENT WEIGHT
Form 2	2A19P. ARRSTD&TKN 2 POL
Form 5	5A21. CURRENT HEIGHT 5A22. CURRENT WEIGHT

## **RECODED VARIABLES:**

### **Core dataset and Part C section of individual forms**

#### **AGE < 18 DICHOTOMY**

1=younger than 18 years old,

2=18 years old or more

-9=missing data on birth year, or birth month if it is required

Derived from Q.C01 (Birth Year), and, if needed, Q.C02 (Birth Month), and the month that the questionnaire was administered. If the birth year value indicates that the respondent is 18, then the month of administration is compared to the month of birth. If the questionnaire was given before the month of birth, or if both were the same month, then the respondent is determined to be younger than 18.

C04. R'S RACE B/W/H -- changed in 2005 from the B/W dichotomy

1=BLACK 2=WHITE 3=HISPANIC, -9=All Other Codes, multiple responses, and missing data on Q. C04.

From 2006 on, each of the questionnaire forms contains the new version of the race question which was introduced on half of the forms in 2005. The new version lists several different response options and prompts the respondents to select all that apply to them. In cases where a respondent selected options which fell into more than one of the three recoded categories (Black, White, Hispanic), the value for the recoded variable was deleted and defined as missing.

#### **C07A. R'S # SIBLINGS**

Responses to questions C07A-B were combined and bracketed before original data were deleted (see above)

0=None, 1=1 sibling, 2=2 siblings, 3=3 or more siblings

#### **C07Cb-d. R'S HSHLD FATHER, MOTHER, SIBLING**

0=marked, 1=not marked, -9=none of the three items marked

#### **C29a-c. # TCKTS AFT [DRNK, MARJ, OTDG]**

0=None, 1=One, 2=Two, 3=Three or More

#### **C31a-c. # ACDTS AFT [DRNK, MARJ, OTDG]**

0=None, 1=One, 2=Two, 3=Three or More

### **Core dataset (Part B)**

\*B10a-c: #X COKE [LIFETIME, LAST12MO, LAST30DA]

Data from forms 1, 3, 4, and 6 are combined responses to separate questions concerning "crack" and "cocaine in any other form".

\*B15a-c: #X "H" [LIFETIME, LAST12MO, LAST30DA]

Data from forms 2, 5, and 6 are combined responses to separate questions concerning heroin "using a needle" and heroin "WITHOUT using a needle".

**Form 6**

A10. EVER HELD BACK

1=No, 2=Yes

A11. NEED SUMMER SCHL

1=No, 2=Yes

A12. EVER SUSPENDED

1=No, 2=Yes

**MISSING DATA FOR WESTERN REGION:**

To ensure confidentiality of both respondents and their respective schools, some variables values from schools in the Western region were changed to missing data (coded -9):

All datasets	C13B R'ATTND REL SVC C13C RLGN IMP R'S LF
Form 2	2A19A FRQ FIGHT PARNTS
Form 4	4A15A FEW GD MAR, ? IT 4A15B GD LIV TG BF MRG 4A15C 1 PRTNR=RSTRCTVE
Form 5	5A18I FAM BUYS THG -ND 5A18J FULLR LVS IF MRY 5A18N HSB MAK IMP DCSN
Form 6	6A08A #X PRNT CHK HMWK 6A08B #X PRNT HLP HMWK 6A08C #X PRNT GV CHORE 6A08D #X PRNT LIMIT TV 6A08E #X PRNT LMT OUT

**QUESTIONNAIRE FORM 1 PROCESSING**

The form 1 questionnaire contains many more specific drug related questions in Part B than do the other questionnaire forms. In the form 1 dataset, copies of the "core" or common drug prevalence variables are created and then processed so that their data will be comparable to that of the other forms. Data from the core versions are then copied to the grade 12 core dataset; the form 1 dataset retains both versions. The primary difference between the copies is that, for the core versions, nonuse is inferred from the respondents' adherence to the skip instructions (the other forms do not include the same instructions).



## **REVISED QUESTION TEXT FOR THE CORE DATASET**

For the core dataset only, additional text was added to particular questions that were part of a series. The initial question in the series contains text not repeated on subsequent questions within that series. This additional text is meant to clarify and provide detail about the question for the user. To help improve the clarity of subsequent questions within the series this additional text has been repeated on each question. This repeating text is identical to what was stated on the questionnaire for the first question in that series. It has been designated by being placed into {braces} to be distinguishable from text that actually appeared in the questionnaire. No modifications were made to the question text for any of the other parts.



# **APPENDIX**



## **Appendix A: Publications**

In previous years, Monitoring the Future Publications were listed as Appendix A to this document.

For a current list of publications referencing Monitoring the Future data, please visit the Monitoring the Future [Publications](#) web page.

Publications are divided into the following categories:

- Monographs
- Reference Volumes
- Books
- Journal Articles
- Chapters
- Research Reports
- Occasional Papers
- Congressional Testimony
- Publications by Study Staff

Many of the publications may be accessed electronically via the web site, either in their entirety and/or in abstract form.



## Appendix B - Sample Size and Student Response Rates

The three-stage sample procedure described in the introduction yielded the following number of participating schools and students.

	<u>Number of Public Schools</u>	<u>Number of Private Schools</u>	<u>Total Number of Schools</u>	<u>Total Number of Students</u>	<u>Student Response Rate*</u>
1975	111	14	125	15,791	78%
1976	108	15	123	16,678	77
1977	108	16	124	18,436	79
1978	111	20	131	18,924	83
1979	111	20	131	16,662	82
1980	107	20	127	16,524	82
1981	109	19	128	18,267	81
1982	116	21	137	18,348	83
1983	112	22	134	16,947	84
1984	117	17	134	16,499	83
1985	115	17	132	16,502	84
1986	113	16	129	15,713	83
1987	117	18	135	16,843	84
1988	113	19	132	16,795	83
1989	111	22	133	17,142	86
1990	114	23	137	15,676	86
1991	117	19	136	15,483	83
1992	120	18	138	16,251	84
1993	121	18	139	16,763	84
1994	119	20	139	15,929	84
1995	120	24	144	15,876	84
1996	118	21	139	14,824	83
1997	125	21	146	15,963	83
1998	124	20	144	15,780	82
1999	124	19	143	14,056	83
2000	116	18	134	13,286	83
2001	117	17	134	13,304	82
2002	102	18	120	13,544	83
2003	103	19	122	15,200	83

	<u>Number of Public Schools</u>	<u>Number of Private Schools</u>	<u>Total Number of Schools</u>	<u>Total Number of Students</u>	<u>Student Response Rate*</u>
2004	109	19	128	15,222	82
2005	108	21	129	15,378	82
2006	116	20	136	14,814	83
2007	111	21	132	15,132	81
2008	103	17	120	14,577	79
2009	106	19	125	14,268	82
2010	104	22	126	15,127	85
2011	110	19	129	14,855	83
2012	107	20	127	14,343	83

\* The student response rate is derived by dividing the attained sample by the target sample (both based on weighted numbers of cases). The target sample is based upon listings provided by schools. Since such listings may fail to take account of recent student attrition, the actual response rate may be slightly underestimated.