

LEARNING FROM EXPERIENCE: ESTIMATING TEEN USE OF ALCOHOL, CIGARETTES, AND MARIJUANA FROM THREE SURVEY PROTOCOLS

FLOYD JACKSON FOWLER, JR., VICKIE L. STRINGFELLOW

Estimates of teen substance use from 3 federally funded surveys were examined. Results from the NHSDA, a household survey, and the MTF and YRBS, both school-based surveys, were compared by gender and ethnic group. Trends in reported use from 1993-1997 were also analyzed. Many discrepancies exist among the results from the 3 surveys and are likely due to several factors, including different sample sizes and frames, policies regarding sample substitution, question wording, and data collection protocols. Discussion of these factors underscores the need for more research dedicated to methodology. Routinely investing a portion of survey budgets in studies of the methods that are used and how they affect comparability with other surveys would build a knowledge base that would make comparisons of results easier and would inform design decisions to increase comparability across surveys.

DEFINING THE PROBLEM

This paper examines the problem of comparing results from three different surveys, all funded by the federal government:

Monitoring the Future (MTF), a survey of 8th, 10th, and 12th grade students which has been carried out annually since 1975 by the University of Michigan. It has been funded by grants from the National Institute on Drug Abuse.

The Youth Risk Behavior Survey (YRBS), a survey of students enrolled in high school (grades 9 through 12) conducted in 1990, 1991, and every other year since then by the Centers for Disease Control (CDC).

The National Household Survey on Drug Abuse (NHSDA), a survey of individuals 12 years old and older living in households, also was initially funded by the National Institute on Drug Abuse and recently has been funded by the

Floyd Jackson Fowler, Jr., Ph.D., is a Senior Research Fellow at the Center for Survey Research, University Massachusetts, Boston. He is the author of several books and numerous papers related to sources of error in surveys. **Vickie L. Stringfellow** is a Senior Assistant Study Director at the Center for Survey Research. Her recent research has focused particularly on instrument design and nonresponse as sources of error in surveys.

Substance Abuse and Mental Health Services Administration (SAMHSA). It was carried out every three years from 1979 to 1991 and annually since then.

Although these surveys have a variety of purposes, they all collect data about substance use, and they include samples of teenagers. It is in those areas of overlap that they provide an opportunity for us to examine the comparability of results from surveys and study the reasons why they produce similar or different results.

ANALYTICAL GOALS

Based on our analyses of these data, there are three different kinds of results of interest.

Point Estimates. Point estimates, estimates of the percentage of some population that has some characteristic or does something. For example, we will be comparing estimates of the rates at which teens use cigarettes. For such estimates, the closer the estimates for the same populations, the better.

Group Comparisons. Group comparisons, such as conclusions about whether males or females are more likely to use cigarettes. Assessment of comparability would be whether or not the data from the surveys produce similar conclusions about how male and female use of cigarettes compares.

Trends. Trends over time. All three surveys provide repeated cross-sectional data designed to measure trends. Another possible comparison of the results across surveys is whether or not they provide similar pictures of how behavior changes over time.

In this paper, we discuss all three of those uses of the data.

APPROACHES TO ANALYSIS

Because the designs of the MTF, YRBS, and NHSDA are different, analyses must be focused on those results that are comparable. Several decisions were made about what to compare.

School Enrollment. MTF and YRBS surveys sample students enrolled in school, while the NHSDA samples people in households. In order to make the results comparable to the school-based samples, we had to restrict the analysis of NHSDA to those respondents who were enrolled in school, whose last completed grade was 8 through 11, and who were 13 to 18 years old. In 1997, the criteria for defining school enrollment were changed. For trending purposes, the 1997 data included separate variables to identify enrollees based on both the original and revised definitions. In order to maintain consistency within the NHSDA sample, we used the enrollment criteria that had been used from 1994 up to and including 1997.

Grade Levels. MTF samples students in the 8th, 10th and 12th grades. The YRBS samples students from the 9th through the 12th grades. In order to achieve comparability between those surveys, the analyses of the three surveys are focused on students in the 10th and 12th grades only.

Substance Use. All the surveys make estimates of the use of numerous substances. The three most used substances by teenagers are cigarettes, alcohol and marijuana. Because the low rates at which other substances are used makes comparing results very complicated, a decision was made to focus on only these three topics.

Current Use. All three surveys gather data about using particular substances during the 30-day period prior to data collection, which is sometimes called "current use." "Current use" seemed perhaps the most salient estimate for monitoring what teens are doing. Our analyses focus on "current use" of the three target substances.

As suggested above, we created tables to make the three kinds of comparisons:

- 1) The rates at which teenagers in the 10th and 12th grades reported current use of the reported substances.

- 2) Analyses comparing groups. As examples, we used the 1997 surveys to make estimates of how males and females compared and how those who considered themselves black, white or Hispanic compared in their current use.

- 3) In order to compare survey results over time, we looked at the estimates from the three surveys of trends of current use for each substance from 1993 through 1997.

To carry out these analyses, we used published data wherever possible. We also performed special-purpose tabulations of NHSDA and YRBS data in order to achieve comparability across the surveys. The sources of our data were the web sites maintained by the Inter-University Consortium for Political and Social Research at the University of Michigan and CDC. The available data sets included weights that are used to adjust for differences in probabilities of selection and for nonresponse. All of our calculations used those weights. All of these surveys have complex designs that are clustered and stratified. We did not feel comfortable taking on the calculation of design effects for our tabulations. Hence, we only calculated statistical tests of differences for the overall 1997 estimates, for which we could use published standard errors plus published design effects for the NHSDA. For the race and gender data, we indicate in the text when differences potentially are and are not statistically significant, but the calculations are only approximations in some cases.

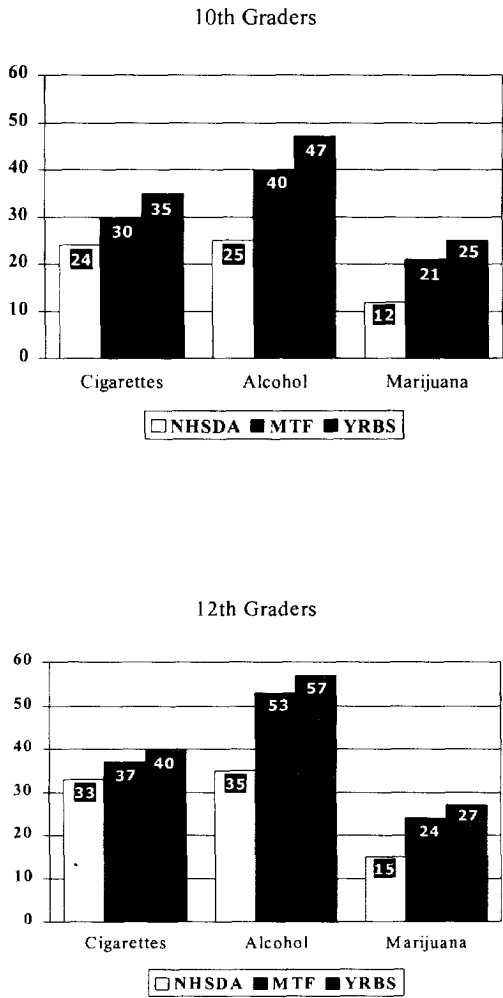
COMPARING RESULTS

Figure 1 presents the results of the first analysis. It shows the reported use in the 30 days prior to data collection of cigarettes, alcohol and marijuana, separately for 10th and 12th graders, based on the data from the three different surveys.

There are several patterns evident from the figure. First, at an order of magnitude level, MTF and YRBS numbers are fairly close, while the NHSDA estimates are notably lower than the other surveys. That pattern is more evident

for alcohol and marijuana than for cigarettes. All 6 estimates in the MTF are significantly higher than the comparable NHSDA estimates, and the same is true for 5 of the 6 YRBS estimates. It is less evident, but it is also true, that YRBS numbers in Figure 1 are without exception higher than estimates from MTF, although some of the numbers are very close. For 3 of the 6 comparisons (all those for the 10th grade), the YRBS estimates are significantly higher ($p < .05$) than the MTF estimates.

FIGURE 1
1997 ESTIMATES OF CURRENT SUBSTANCE USE FOR 10TH AND 12TH GRADERS BY SOURCE



There are some similarities in the numbers from the three surveys that are worth noting. In all cases, current use is estimated to be higher for 12th graders than among 10th graders. In addition, all surveys indicated that alcohol use was higher than cigarette use, which in turn was higher than marijuana use. With respect to the latter generalization, however, the estimates from the NHSDA are very similar regarding use of cigarettes and alcohol, while the ordering is much clearer from MTF and YRBS data.

The next two figures look in more detail at relationships. The first compares reported 30-day substance use by whites, blacks and Hispanics; the subsequent figure compares reported male and female use of these substances.

Starting with ethnic background, the three surveys tell a fairly similar story with respect to use of cigarettes and alcohol. From all three surveys, one would conclude black teens were less likely than white teens to use either cigarettes or alcohol. In 5 of the 6 comparisons for cigarette use in Figure 2, Hispanics are somewhere between white and black teens. For alcohol, Hispanics consistently report more use than blacks, and to varying degrees approach the rates reported by whites.

The apparent consistency among surveys shown for alcohol and cigarette use is not replicated with respect to marijuana. Two of the surveys, MTF and the NHSDA, show black teens using marijuana less than white teens. Using average design effect estimates, the difference between black and white 10th graders in the NHSDA is almost certainly statistically significant. In contrast, the YRBS survey indicates that, if anything, black teens use marijuana more than whites in both the 10th and 12th grades. The figures with respect to marijuana use by Hispanics vary from survey to survey in comparison to their white and black peers.

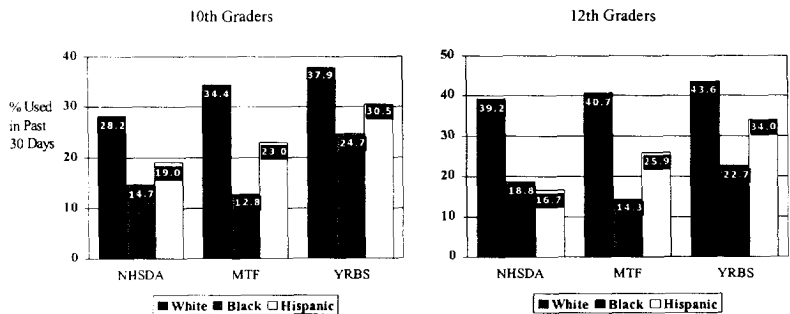
Figure 3 produces a similar analysis for males and females in the 10th and 12th grades. In this case, the survey results all show males and females to be quite similar with respect to cigarette use in both the 10th and 12th grades. They are also similar in their reported alcohol use in the 10th grade, but males pull ahead of the females by the 12th grade. All three surveys show that males use marijuana more than females in both the 10th and 12th grades. Most of the male-female differences in 12th grade alcohol use and for marijuana use are statistically significant ($p < .05$).

The graphs in Figure 4 map trends over time, from 1993 to 1997. Note that there are two comparability issues. While the MTF collected data every year in a manner thought to be consistent, the YRBS collected data only every other year (1993, 1995, and 1997). The household survey collected data every year, but it changed procedures between 1993 and 1994, rendering the 1994 and more recent estimates incomparable to prior estimates. As a result, data in the table are only mapped from 1994 on for the NHSDA.

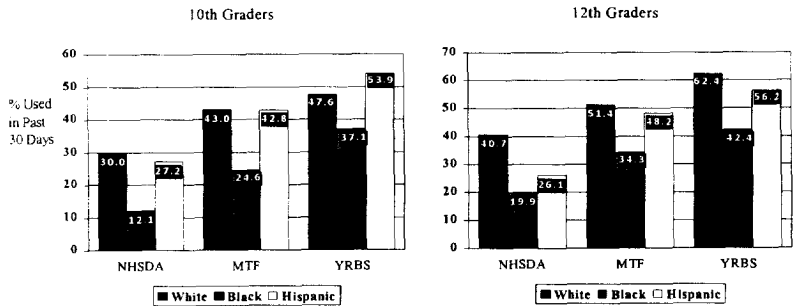
The graphs map trends over time for the 10 and the 12th graders, separately for each of the three substances. While the shape and slope of the trends vary some from survey to survey, all three surveys show a tendency for cigarette use to be up from 1993 to 1997 for both 10th and 12th graders. The alcohol data are less consistent. The trends from the two school-based surveys, YRBS and MTF, look

FIGURE 2
1997 ESTIMATES OF CURRENT CIGARETTE, ALCOHOL, AND MARIJUANA USE
BY GRADE AND ETHNICITY

Cigarette Use



Alcohol Use



Marijuana Use

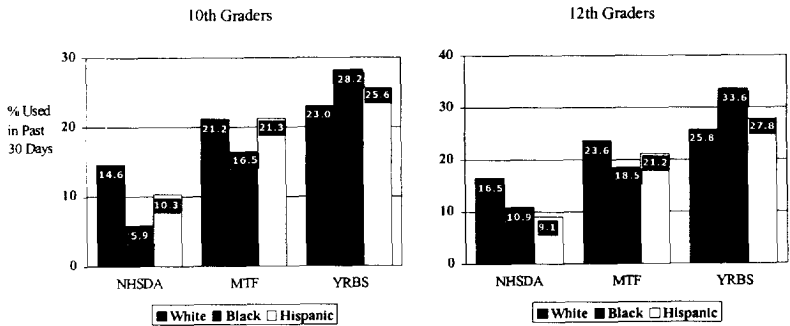
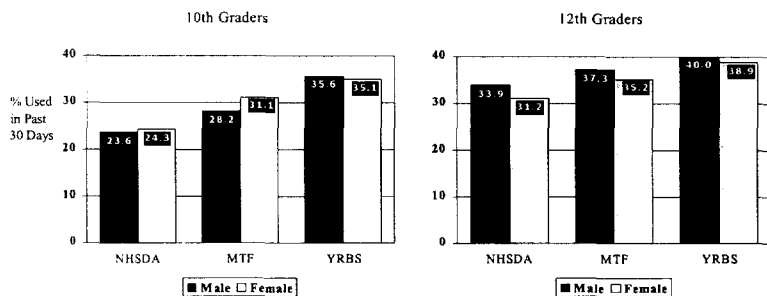
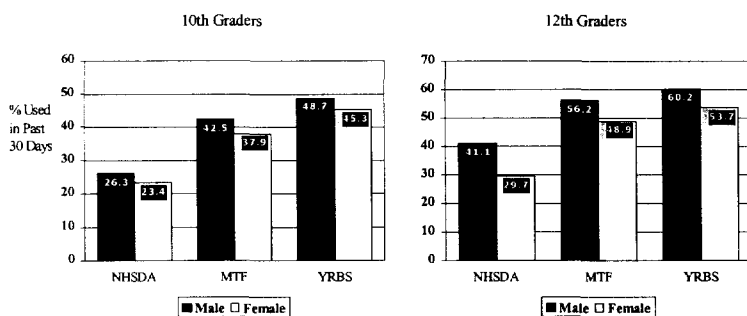


FIGURE 3
1997 ESTIMATES OF CURRENT CIGARETTE, ALCOHOL, AND MARIJUANA USE
BY GRADE AND GENDER

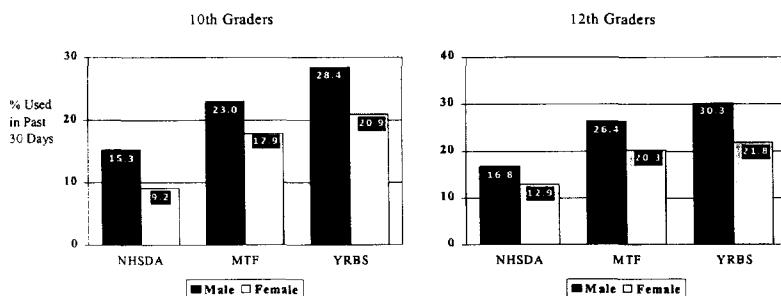
Cigarette Use



Alcohol Use



Marijuana Use



as if alcohol use was nearly flat from 1993 through 1997. In contrast, based on the NHSDA, for both grades, the 1997 estimate is lower than the 1993 estimate. Thus, the NHSDA suggests that alcohol use declined among 10th and 12th graders. Both the school-based surveys indicate that marijuana use rose between 1993 and 1997, though the YRBS gets the same estimates from 1995 to 1997. The NHSDA also indicates a rise in marijuana use among 10th graders from 1994 through 1997. The NHSDA, in contrast to the other surveys, reports that marijuana use among 12th graders was quite stable from 1994 to 1997. Yet, to compare results a different way, if we look only at the period 1995 through 1997, all three surveys, for both 10th and 12th grades, indicated that use was fairly stable.

One other point about the trend data is worth noting. The 1997 data in Figure 1 showed that there was a consistent ordering among the surveys, with the YRBS having the highest estimates and the NHSDA having the lowest estimates of use for all three substances. It is apparent from the lines in Figure 4 that 1997 was not an anomaly. Without exception, in every graph, all of the estimates from the YRBS are higher than any of the other two estimates; all of the NHSDA estimates are lower than those from the other two surveys. Thus there is compelling evidence that there is something about the survey protocols that produces consistently different levels of use over time for 10th and 12th graders and for all three substances.

In conclusion, we might summarize the comparability as follows:

- 1) There is a consistent ordering across the surveys in rates at which teens are estimated to use these substances;
- 2) When comparing white, black and Hispanic teens, the patterns are quite similar for cigarettes and alcohol, but the patterns projected for marijuana differ across surveys. YRBS is the outlier. The relationship between gender and use is consistent for all three surveys for all three substances.
- 3) With respect to trends over time, there also was considerable consistency, but there were three (out of 6) graphs for which the NHSDA trend looked somewhat different from the school-based surveys.

SOURCES OF ERROR IN SURVEYS

A survey research project is actually a melding of three methodologies: sampling, data collection and question design. How well survey results provide accurate descriptions of some population depends upon how each of these aspects of the survey research process is carried out.

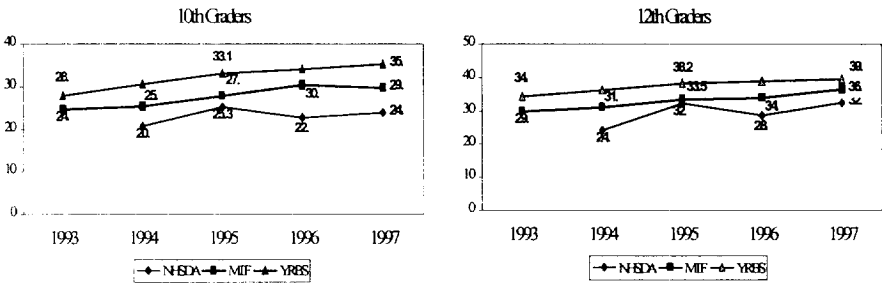
If the people from whom data are collected are not a good representation of the whole population, describing the sample will not do a good job of describing the population. In addition, the quality of measurement depends critically on how the data are collected and how the questions are designed.

We will consider several possible sources of difference among the three surveys in the following discussion, including sample size and design, the sample

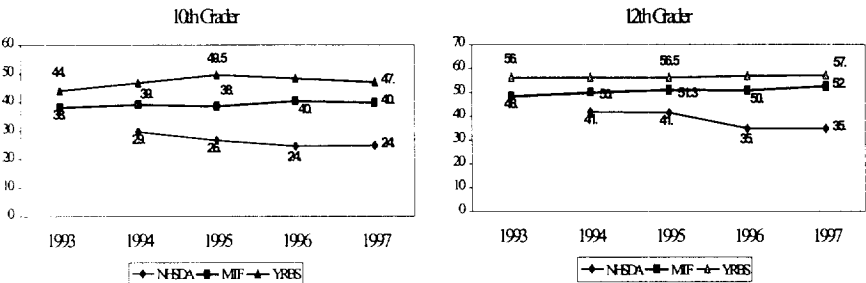
frames, nonresponse, data collection protocols, and the content of the survey instruments.

FIGURE 4
ESTIMATES OF CURRENT CIGARETTE, ALCOHOL, AND MARIJUANA USE
BY GRADE AND YEAR COMPARING SAMPLES

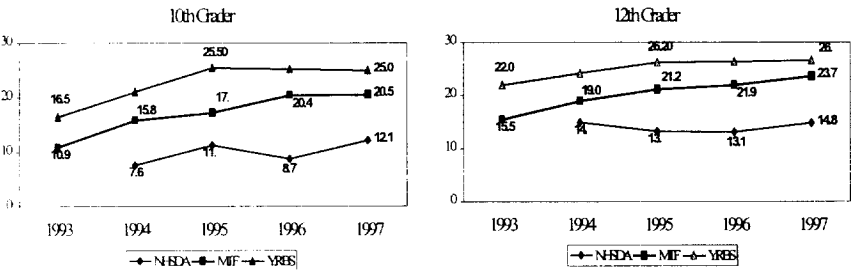
Cigarette Use



Alcohol Use



Marijuana Use



COMPARING SAMPLES

Survey samples differ in size, who is in the sample frame and has a chance of being selected, and who among those selected actually ends up responding.

TABLE 1
1997 SAMPLE SIZES* OVERALL AND FOR SUBGROUPS BY SURVEY

	<u>NHSDA</u>		<u>MTF</u>		<u>YRBS</u>	
	<u>10th</u>	<u>12th</u>	<u>10th</u>	<u>12th</u>	<u>10th</u>	<u>12th</u>
Total	1041	741	15500	15400	3858	4386
Male	502	318	7400	7100	1958	2080
Female	539	423	7800	7700	1932	2329
White	518	477	20900**	19800**	1355	1389
Black	194	132	3200**	3600**	1005	1290
Hispanic	262	180	3200**	2800**	1128	1278

*All numbers are unweighted. Some are rounded. Numbers on which estimates are based may vary due to item non-response. The effective sample sizes often are much smaller due to the effects of clustering and oversampling of population subgroups.

**Includes responses from 1996 and 1997

SAMPLE SIZE

Table 1 presents the 1997 sample sizes for these surveys for the data presented earlier in this paper. It is evident that the sample sizes vary greatly. The MTF samples, designed to produce grade-specific analyses, are by far the largest. The NHSDA, which collects data for all ages over 12, has the smallest samples of the 10th and 12th graders. Estimates for black and Hispanic teens are based on particularly small samples for the NHSDA.

When samples are small, as they are for the NHSDA, there is increased likelihood that some estimates will differ from true values by chance alone. However, the most striking difference among the surveys when comparing groups was the reported use of marijuana by different ethnic groups. In that case, it was the YRBS that was deviant; the NHSDA and MTF surveys showed very similar results.

SAMPLE FRAMES

The sample frames for the MTF and YRBS surveys are very similar. Multi-stage samples, first drawing geographic areas, then sampling schools, then sampling classes within schools, was the approach used for both surveys.

Both surveys routinely experience a good bit of non-response. The YRBS school response rate was 79% in 1997. In contrast, in 1997 only 50% of the MTF

schools initially selected agreed to participate. The MTF survey routinely experiences a higher rate of schools refusing to participate than the YRBS.

For both surveys, there is further non-response due to students in selected schools not completing a survey. The most important source of non-response of that sort comes from absenteeism, students who are not in school on the day data are collected. A very small number refuse. In 1997, 89% of students in participating schools responded to the YRBS; the comparable rates for the MTF study were 86% for 10th graders and 83% for 12th graders. The percentage of the population reflected in the returns, of course, is the product of the school response rate and the rates at which students in participating schools responded.

The NHSDA is based on a probability sample of housing units. Individuals associated with housing units are listed, then sampled. Most years, over 80% of sampled individuals complete an interview.

There are several sample-related differences among the three surveys that might affect the answers.

Inclusion of Dropouts. The most obvious difference among the samples is that the household survey includes teens who have dropped out of high school. It is documented that dropouts use illicit substances more than those who stay in school (Gfroerer, Wright, & Kopstein, 1997). To make estimates comparable, those in the household survey not enrolled in school can be eliminated from tabulations (as they have been above). For those interviewed in the summer, there is a slight ambiguity about whether or not they are currently enrolled in school; a few may consider themselves enrolled during the summer but not re-enroll in the fall (or vice versa). However, analyses (Gfroerer et al., 1997) suggest that it is not an issue that affects estimates very much.

School Nonresponse. The second important difference between the MTF and the YRBS is how they handle school non-response. The MTF survey substitutes "similar" schools when a school declines, while the YRBS does not. The MTF staff presents arguments for why the substitution process does not affect estimates in a systematic way (Bachman, Johnston, & O'Malley, 1996). They argue that schools are not highly related to the prevalence of use of the various substances. However, it is fair to say that matching schools within a geographic area is an imprecise science. At the very least, there is the potential for some kind of unknown and unpredictable effect on any year's estimates that would not be present in the other two surveys. The question of whether or not the substitution process could be responsible for the slightly lower estimates of use in the MTF compared with the YRBS is difficult to evaluate. It certainly is one possibility to explain that difference.

Individual Nonresponse. All three surveys run the potential risk of having individual non-respondents be systematically more likely to be substance users than respondents. Studies of the school-based samples suggest that those who are absent, who constitute almost all of the non-respondents, are more likely to be using substances than those who are present (Johnston, O'Malley, & Bachman,

1998). Although harder to evaluate, it also seems plausible that individuals in households who use drugs may be more difficult to schedule for an interview. The fact that for the school-based samples, substance users who show up at school on the day the survey is done have very little chance to avoid filling out a survey may or may not be an advantage. Altogether, it is hard to assess how individual non-response might differentially affect the estimates from the three surveys. The rates of individual non-response, once schools have agreed to participate, are rather similar across all surveys.

Weighting. The potential to adjust or weight responses to make results more comparable seems quite minimal. Because substance use is widely distributed, and not highly correlated with any easily identified characteristic, there are no clear adjustments that would improve estimates of any of the surveys. University of Michigan researchers report on various attempts to adjust for factors that might affect estimates, and the effect of those efforts is consistently small (e.g. Johnston et al., 1998). It is true that gender and ethnic background are systematically related to substance use in predictable ways. However, the potential to substantially affect aggregate estimates by such adjustments is modest.

In conclusion, other than focusing on the school-based population from the NHSDA, which is necessary to make results from the three surveys comparable, and the fact that smaller samples are subject to more sampling error, there really are only three issues that one can see with respect to sampling that might lead to differences between surveys: the non-response at the school level that the MTF and YRBS surveys experience, the fact that MTF substitutes schools for non-responding schools in ways that cannot be totally matched, and concern about who the individual non-respondents are in households or in schools. It is acknowledged that the individual non-respondents are more likely to be substance users, and hence that element of non-response probably leads to slight underestimates of substance use in all three surveys. It is not clear that the effect would differentially affect the surveys. The potential that the MTF substitute schools are slightly lower with respect to drug use than the initially sampled schools remains a plausible idea that cannot be ruled out. It may be part of the explanation for why the MTF estimates are lower than the YRBS estimates on a routine basis.

COMPARING DATA COLLECTION PROTOCOLS

It is well established in the literature that people tend to underreport behaviors or characteristics that may be negatively judged. Survey strategies that are viewed as more private and make respondents feel less vulnerable to judgement tend to improve reporting. Since the use of cigarettes, alcohol, and marijuana by teens constitute behaviors that most teens would not like to be public, at least to the adult world, there is every reason to think that the degree of privacy and protection that respondents feel has a marked effect on the rates at which substance use is reported.

The three surveys differ substantially in their data collection protocols. The household survey involves an interviewer coming to the respondent's home. The answers are recorded by the respondent on self-administered sheets. Earlier studies by NIDA demonstrated that self-administered reports were higher than those provided directly to interviewers (Aquilino & Losciuto, 1990; Turner et al., 1992). Special studies have shown, however, that a significant portion of the time interviewers are involved to some extent in helping respondents with the forms, and hence interviewers are able to see the answers that are given. In addition, it is not uncommon for parents or other relatives to be nearby when the data collection is taking place.

In contrast, the school-based surveys are completed in a group setting. Questionnaires without names on them are dropped into a box; there is no reason to think that any adult could see any of the answers (though it is possible that peers could be peeking at answers). Also, the presence of a teacher nearby could be inhibiting. One difference between the MTF and YRBS studies is the following: The MTF has had a component that involves collecting follow-up data from respondents at a later date. As a result, questionnaires are identified with a serial number, and respondents separately provide their names and addresses that can be linked to that number. This procedure enables researchers to contact the students at a later date and to link later answers to the original responses. Thus, MTF respondents know they can be linked to their answers, while that is not the case for the YRBS respondents.

There seems almost no doubt that the reason for the consistently lower estimates that come from the NHSDA is the mode of data collection. In a 1997 experiment, the standard protocol was compared with an audio computer-assisted self-interviewing procedure (ACASI) (Penne et al., 1998). In connection with the experiment, respondents were asked questions about the interviewing environment, privacy and the like. Focusing just on the 12 to 17 year old respondents, for those respondents whose data collection protocol mirrored those used to collect the data reported in this paper, a third reported getting some help from interviewers in answering the questions, only 56% said they felt "very comfortable" using the answer sheets to answer questions about cigarettes and alcohol, and, perhaps most interestingly, over 60% of the respondents said the interviewer saw at least some of the answers that they had marked. In all these respects, the computer-assisted strategy provided much more privacy. When asked whether the computer or the written answers sheets were best for protecting privacy, almost half picked the computer, while only 10% chose the answer sheets.

That experiment went on to show that the computer-assisted protocol produced a marked increase in the reports of use of various substances, primarily among teens 12 to 17. The increased reporting was less apparent in reports of cigarette and alcohol use than for some other drugs. Thus, that study alone does not guarantee that this is the sole problem affecting the cigarette, alcohol, and

marijuana comparisons. Nonetheless, both the previous studies on this topic and the information about how teens experience the NHSDA interview make for a very strong case that the data collection protocols themselves play a big role in the NHSDA getting lower estimates of substance use than the school-based surveys.

With respect to the differences between the two school-based surveys, it also seemed quite reasonable to think that the different confidentiality protocols might be the cause of the consistent differences between the MTF and YRBS data. Remember, the MTF survey consistently has collected data to enable them to recontact respondents. Hence, MTF respondents knew they were identifiable, albeit only by a coded number and with a promise of confidentiality. The notion that such a difference might produce a few percentage points difference in the rates at which people reported substance use seemed plausible. However, researchers at the University of Michigan recently discontinued doing follow-up surveys of students in the 8th and 10th grades. They conducted an experiment in which 8th and 10th graders were randomized either to a confidential or an anonymous condition (O'Malley, Johnston, & Bachman, 2000). The finding was that there was no discernible effect on their reporting of use of various substances. Thus, while the lack of anonymity in the MTF seems an obvious candidate for explaining some of the differences between it and the YRBS, this experiment makes that explanation less likely to be true.

One final issue should be addressed: The potential impact of filling out an instrument in a school context surrounded by peers. People have raised the possibility that such an environment could stimulate overreporting by students, either because they would not take the survey seriously or because presenting a profile to peers of someone who does a lot of risky things might be attractive. The bottom line implication of such an analysis is that while the household survey may have consistent underreporting due to concerns about privacy, school-based surveys may have overreporting due to some kind of perverse pressures from the environment and from peers. This hypothesis is very difficult to assess. CDC did a test-retest study of reporting and found that the answers were reasonably stable over a two-week period (Brener et al., 1995). The prevalence estimates were also virtually identical. While that does not prove that the answers were not distorted in some way on both occasions, the burden of remembering an exaggerated answer is harder than remembering a true one. All in all, it is very difficult to assess definitively whether student needs to be entertaining or to impress peers could in some cases have led to distorted reporting.

In conclusion, it seems likely that the most important source of divergence among the three surveys is the data collection protocols. Specifically, the perceived lack of privacy involved in the household data collection almost certainly has a dampening effect on teens' reports of use of various substances and is largely responsible for the fact that lower rates of use are obtained in the

NHSDA than in the school-based surveys. Whether or not some overreporting in the school environment exacerbates the differences is very hard to assess.

COMPARING QUESTIONS AND ANSWERS

Tables 2 and 3 lay out the question series for the three surveys for measuring 30-day use of cigarettes, alcohol and marijuana.

CIGARETTES

The three survey instruments differ in how specific they are about what counts as cigarette use. The MTF instrument simply says, "smokes cigarettes," without elaborating further. In contrast, the NHSDA instrument specifically says, "number of days I smoked at least a puff or two from a cigarette." Thus, the NHSDA is explicit that one need not smoke a full cigarette in order to count as a day of cigarette use. The YRBS is a bit of a mixture. In two preceding questions, it specifies that for ever smoking, "even one or two puffs" counts. On the other hand, in the next question, the respondent had to smoke a whole cigarette to report when he or she smoked for the first time. Which of those standards applies to the 30-day question is not specified; all that is said is, "How many days did you smoke cigarettes?"

Of the three instruments, one would expect that the NHSDA questions would produce the highest number of days in which someone smoked in the past 30 days because they specify that a partial cigarette counts. The two school-based surveys are probably similar, in that both essentially leave it to the respondent to decide what counts as smoking a cigarette.

The response tasks for the three surveys also varied in ways that could affect the rates at which people reported current smoking. The MTF has two features that might produce lower reports than the other two surveys. First, the MTF is the only survey that has a screening question. If the respondent says that he or she never smoked cigarettes, there is a skip over the question asking about use of cigarettes in the past 30 days. In contrast, everyone must answer the 30-day smoking questions in the other two instruments. There is some research that suggests that such screening questions may reduce the rate at which people report doing something. On the other hand, an experiment with the NHSDA directly compared asking a screening question with a skip to asking the unscreened version of the substance questions. In that test, no differences were found (Penne et al., 1998). Hence, the screening question may not be responsible for any difference in the results.

Another difference among the three instruments is the form of the response task. The NHSDA asks respondents to fill in the number of days on which they smoked a cigarette, without using response alternatives. The YRBS asks for the number of days in the past 30 days in ordered categories. The MTF asks for number of cigarettes per day. The lowest category above "not at all" is "less than one cigarette per day." This is an odd response to choose for someone who has

smoked cigarettes on only a few days in the past month. It would not be unreasonable to think that some respondents who smoked only occasionally would choose the “not at all” response as being closer to the truth. Again, from a questionnaire design point of view, one would say that the NHSDA is the best design for getting large numbers; it has been shown that having people answer an open ended form increases the number of socially undesirable events or occasions reported (Sudman & Bradburn, 1982); one would guess that the MTF approach to the question would be most likely to produce a lower estimate of cigarette use.

TABLE 2
QUESTIONS ABOUT CURRENT CIGARETTE USE BY SURVEY

NHSDA	MTF	YRBS
How long has it been since you last smoked a cigarette? - Within the last 30 days - More than 30 days ago but within the past 12 months - More than 12 months ago but within the past 3 years - I have never smoked a cigarette in my life	1. Have you ever smoked cigarettes? - Never – Go to Question 3 - Once or twice - Occasionally but not regularly - Regularly in the past - Regularly now	
THE NEXT QUESTIONS REFER TO THE PAST 30 DAYS ONLY Think specifically about the past 30 days – that is, from your 30-day reference date up to and including today. During the past 30 days, on how many days did you smoke a cigarette? - Number of days I smoked at least a puff or two from a cigarette - I have smoked cigarettes but not during the past 30 days - I have never smoked a cigarette in my life	2. How frequently have you smoked cigarettes during the past 30 days? - Not at all - Less than one cigarette per day - One to five cigarettes per day - About one-half pack per day - About one pack per day - About one and one-half packs per day - Two packs or more per day	During the past 30 days, on how many days did you smoke cigarettes? - 0 days - 1 or 2 days - 3 to 5 days - 6 to 9 days - 10 to 19 days - 20 to 29 days - All 30 days

ALCOHOL AND MARIJUANA

Turning to the questions about use of alcohol, the issues are fairly similar. In this case, all three surveys tell respondents what kinds of beverages count as alcoholic beverages, and they all say that just a few sips does not count as a drink. (There is an interesting variation in how they modify sips. The NHSDA says sips from another person's drink do not count; the YRBS says that sips for religious

purposes do not count; the MTF says that a few sips do not count, no matter where they came from.) Of the three, only the NHSDA tries to define the amount of each alcoholic beverage that counts as "a drink." The school-based surveys are both silent on that issue.

Once again, the MTF has a screening question. If respondents answer, "no," that they have never had more than a few sips of alcohol, they skip the question about frequency in the last 30 days. There are no skips in the other two instruments.

Overall, while there are differences in how they specify what is counted, and generally the NHSDA wording probably is the most specific, it is not clear that these differences in wording would systematically produce differences in the estimates of frequency of current use.

The three surveys also differ in the response task they offer. Again, the NHSDA asks respondents to give the number of days in the last 30 days in an open-ended fashion; the YRBS and MTF surveys use a very similar set of ordered categories of numbers for response tasks. However, there are two distinctive features of the MTF questions that might affect the results. First, MTF specifies, "occasions" in the last 30 days on which you have used alcoholic beverages while the other two surveys ask for the number of days. Presumably, that would allow for the possibility of someone reporting multiple occasions on the same day. That may not be a common event; if it happened, it would tend to increase the number of "occasions" reported in the MTF.

A likely more important feature of the MTF instrument is the way it asks for use of alcohol, "in your lifetime," "during the last twelve months," and "during the last 30 days." This kind of question structure focuses respondents on the reference period. Based on some studies by Loftus, Smith, Klinger, and Fiedler (1991), one might expect this kind of design to result in less 30-day use reported, and that the reporting might be more accurate, than when the 30-day question stands alone. The Loftus research shows that when people are asked to report the frequency of events in a particular reference period, if they have not done what is being asked exactly within the reference period but have done it recently, they are inclined to report it anyway, thinking it is a more accurate answer. By giving people a chance to report their annual alcohol use, the question highlights the meaning of the 30-day reference period and may make respondents feel that they are not compelled to report 30-day use unless it is literally true.

Turning to marijuana use, the issues are virtually identical to the alcohol use questions, except there is not a screening question in the MTF. Once again, the NHSDA survey stands out for providing more information about what is meant by marijuana, though all three instruments have some alternative wording. The NHSDA also makes it clear that marijuana can be used in numerous ways, not simply smoked. The NHSDA respondents answer in open ended form, while the two other surveys have a set of ordered categories. The unit varies across the surveys: NHSDA uses "days," MTF uses "occasions," and YRBS uses "times."

Once again, the MTF question is embedded in a series that asks about three time periods, with 30-day use being last. In this case, the YRBS also asks a lifetime use question prior to the 30-day use question.

TABLE 3
QUESTIONS ABOUT CURRENT ALCOHOL AND MARIJUANA USE BY SURVEY

NHSDA	MTF	YRBS														
Think about the last time you drank any type of alcoholic beverage. How long has it been since you last (drank an alcoholic beverage/used marijuana or hashish)? - Within the past 30 days - More than 30 days ago but within the past 12 months - More than 12 months ago but within the past 3 years - More than 3 years ago - I have never (drunk an alcoholic beverage/used marijuana or hashish) in my life	<i>(No screener for marijuana use.)</i> Next we want to ask you about drinking alcoholic beverages, including beer, wine, wine coolers, and liquor. Have you ever had any beer, wine, wine coolers, or liquor to drink – more than just a few sips? - No – Go to Top of next Column - Yes															
Think specifically about the past 30 days -- that is, from your 30-day reference date up to and including today. During the past 30 days, on how many days did you (drink one or more drinks of alcoholic beverages/use marijuana or hashish)? - Number of days I (had a drink of an alcoholic beverage/used marijuana or hashish) - I have (drunk alcoholic beverages/used marijuana or hashish) but not during the past 30 days - I have never (drunk an alcoholic beverage/used marijuana) in my life	On how many occasions have you (had alcoholic beverages to drink – more than just a few sips/used marijuana [weed, pot] or hashish [hash, hash oil])? <u>Occasions for each item</u> <table><tr><td>...in your lifetime?</td><td>0 occasions</td></tr><tr><td>...during the last 12 months?</td><td>1-2 occasions</td></tr><tr><td>...during the last 30 days?</td><td>3-5 occasions</td></tr><tr><td></td><td>6-9 occasions</td></tr><tr><td></td><td>10-19 occasions</td></tr><tr><td></td><td>20-39 occasions</td></tr><tr><td></td><td>40 or more times</td></tr></table>	...in your lifetime?	0 occasions	...during the last 12 months?	1-2 occasions	...during the last 30 days?	3-5 occasions		6-9 occasions		10-19 occasions		20-39 occasions		40 or more times	During the past 30 days, on how many days did you (have at least one drink of alcohol/use marijuana)? - 0 days - 1 or 2 days - 3 to 5 days - 6 to 9 days - 10 to 19 days - 20 to 29 days - All 30 days
...in your lifetime?	0 occasions															
...during the last 12 months?	1-2 occasions															
...during the last 30 days?	3-5 occasions															
	6-9 occasions															
	10-19 occasions															
	20-39 occasions															
	40 or more times															

Overall, there are substantial differences in the way these questions are asked, differences that could well affect the estimates. In general, the NHSDA questions are the ones that would be expected to produce the highest and best estimates on question form alone, because they are more specific about what to count and they allowed people to answer in open-ended form. On the other hand, there are two features of the MTF instrument design that may help to account for why their estimates are consistently lower than the YRBS. For cigarettes, the response task,

which asks for number of cigarettes per day in the last 30-days, seems potentially to lend itself to increasing false negatives, the number of people who say, "no days" who in fact smoke cigarettes at least occasionally. On the other hand, the design of the questions for alcohol use and marijuana use, by having the 30-day questions follow lifetime and annual questions, is a way of improving the quality of reporting and actually reducing overreporting.

QUESTION CONTEXT

The context, the position that questions occupy in the survey instrument, sometimes can affect the way questions are interpreted and answered. The content of the three instruments is quite different. The NHSDA survey is almost totally devoted to substance use. The YRBS, in contrast, focuses on many behaviors and issues that might affect health. Suicide, pregnancy, and fighting are all explored in the instrument, as are eating vegetables and getting enough exercise. The MTF instrument asks about respondents' values and plans for the future.

It is hard to estimate how the differences in overall survey content might affect respondents' willingness to accurately report the use of cigarettes, alcohol, and marijuana. It is possible that context has two effects on reporting. First, if cigarettes, alcohol, and marijuana use were the most serious things one could report in an instrument, it is possible that respondents would be more reluctant to admit to use than if there were other worse things in the instrument to deal with. On a related point, people are more willing to divulge some less than perfect characteristics of themselves if they also are allowed to say positive things about themselves. Thus, an instrument in which respondents are able to say they are doing some positive things at the same time they are admitting to doing some things that are not highly desirable might create an environment in which people would be more willing to report. This kind of thinking is consistent with some of the work that Sudman has done on this topic (e.g. Sudman & Bradburn, 1982).

Following that line of thought, all three of the instruments ask about much worse substances than cigarettes and marijuana, but the MTF and YRBS offer more opportunities for respondents to report on positive behaviors than does the NHSDA, which focuses almost exclusively on negative behaviors. Thus, the context may help to explain why there is more reporting of use in the two school-based surveys than in the household survey.

LESSONS TO BE LEARNED

The core lessons from the analyses in this paper are that it is very difficult to compare results from independent surveys and it is difficult to reach firm conclusions about why survey results differ, when they do differ. There are three reasons for this.

DISSIMILAR STATISTICS

Unless a special effort is made to create statistics that are strictly comparable, results as reported from different surveys are unlikely to be strictly comparable.

A good deal of data manipulation was necessary to produce the tables and comparisons in this paper. Despite the fact that these data sets are extensively analyzed, with analyses largely focusing on the teen population and on similar measurement objectives, without special tabulations it would be virtually impossible to strictly compare the results.

NUMEROUS DIFFERENCES IN METHODS

Because survey estimates are affected by so many factors, unless very special effort is made to conduct surveys that are comparable, surveys are likely to differ in numerous ways that could potentially affect survey estimates. As a result, when survey estimates differ, it may not be easy to decide which estimate is best. For example, of the three surveys compared here, the NHSDA probably has the stronger methodology in at least three respects: the sample frame is more comprehensive (for instance including people in and out of school), the overall response rate is higher, and in a number of ways the survey question design could be argued to be stronger. Nonetheless, it is probable that the survey's lower estimates reflect response error, brought on by the less private data collection protocol.

Meanwhile, the YRBS and MTF surveys systematically differ year after year, with the MTF estimates of use being slightly lower. While they share a very similar data collection protocol, they differ in response rate, in how they handle school nonresponse, and in some question wording issues. While they produce similar conclusions on a number of issues we looked at, when they do diverge, it is difficult to say which one is most likely to be accurate.

LACK OF GOOD STUDIES

The third reason it is difficult to reach firm conclusions about which survey estimates are best is there has not been enough invested in studies directed specifically at methodology. As a result, when survey methods differ, it often is difficult to know which differences are likely to affect data, and which of the variations used is likely to be best.

Two studies that were extremely helpful in sorting out the issues were not planned to reconcile differences among these three surveys. The NHSDA study that compared previous practice to a computer-assisted data collection protocol also collected data about teen sense of privacy when filling out the self-administered forms. Those data greatly strengthened the hypothesis that a key difference between the household survey and the school-based surveys was the sense of privacy and comfort that students would feel in reporting substance use. When the University of Michigan researchers decided not to do follow up surveys of 8th and 10th graders, their wisdom in randomizing students to an anonymous and a confidential condition provided important evidence that the lack of anonymity *per se* probably did not affect substance use reporting. In other areas

in which the surveys differ, however, we lack good methodological research that would help us assess the significance of the differences.

CONCLUSION

More research on methods is needed. As we have seen, there are many ways in which methodology can affect the final results of a study. More information about the implications of factors such as nonresponse, question wording, and mode of data collection is necessary to make informed decisions when designing and executing surveys.

Furthermore, comparing results from disparate surveys will continue to be a challenge. However, if there were a commitment to invest routinely a small portion of survey budgets in studies of the methods that are used, to assess the significance of some of the decisions that are made and how they affect comparability with other decisions that might have been made, we would build a knowledge base that would make comparisons of survey results easier. Moreover, that knowledge base would greatly inform the decisions that are made in the design of future surveys.

ACKNOWLEDGMENTS

This paper was prepared under a contract with Westat. The work was sponsored by the Office of the Assistant Secretary for Planning and Evaluation, U. S. Department of Health and Human Services.

REFERENCES

- Aquilino, W.S., & Losciuto, L.A.
1990 Effects of interview on self-reported drug use. *Public Opinion Quarterly*, 54, 362 - 391.
- Bachman, J.G., Johnston, L.D., & O'Malley, P.M.
1996 *The monitoring the future project after twenty-two years: Design and procedures*. Ann Arbor, MI: Institute for Social Research (Monitoring the Future Occasional Paper, 38).
- Brener, M.D., Collins, J.L., Kahn, L., Warren, C.W., & Williams, B.I.
1995 Reliability of the Youth Risk Behavior Survey Questionnaire. *American Journal of Epidemiology*, 141(6), 575-580.
- Gfroerer, J., Wright, D., Kopstein, A.
1997 Prevalence of youth substance use: The impact of methodological differences between two national surveys. *Drug and Alcohol Dependence*, 47, 19-30.
- Johnston, L.D., O'Malley, P.M., & Bachman, J.G.
1998 *National survey results on drug use from the Monitoring the Future Project, 1975-1997*. Rockville, MD: National Institute on Drug Abuse.

- Loftus, E.F., Smith, K.D., Klinger, M.R., & Fiedler, J.
1991 Memory and mis-memory for health events. In J. Tanur (Ed.), *Questions about Questions: Inquiries into the cognitive basis of surveys* (pp. 102-137). New York: Russell Sage Foundation.
- O'Malley, P.M., Johnston, L.D., Bachman, J.G., & Scholenberg, J.
2000 A comparison of confidential versus anonymous survey procedures: Effects on reporting of drug use and related attitudes and beliefs in a national study of students. *Journal of Drug Issues*, 30(1), 35-54.
- Penne, M.A., Lessler, J.T., Bieler, G., & Casper, R.
1998 Effects of experimental audio computer-assisted self-interviewing (ACASI) procedures on reported drug use in the NHSDA: Results from the 1997 CAI Field Experiment. *Proceedings of the section on survey research methods, American Statistical Association*, 744-749.
- Sudman, S., and Bradburn, N.
1982 *Asking Questions*. San Francisco: Jossey-Bass.
- Turner C, Lessler, J., DeVore, S.
1992 Effect of mode of administration and wording on reporting drug use. In C.F. Turner, J.T. Lessler, & J.A. Gfroerer (Eds.), *Survey Measurement of Drug Use* (pp.177-219). Washington, DC: US. Government Printing Office.