Agreement Between Retrospective Accounts of Substance Use and Earlier Reported Substance Use

Linda M. Collins, John W. Graham, William B. Hansen, and C. Anderson Johnson
University of Southern California

This present study examined agreement between retrospective accounts of substance use and earlier reported substance use in a high school age sample. Three issues were addressed: (1) extent of overall agreement; (2) evidence for the presence of a response-shift bias; and (3) extent to which current use biases recall of substance use. Subjects were 415 high school students who took part in a smoking prevention program. At the last measurement, which took place 2½ years after the pretest, the students were asked to recall pretest use of tobacco, alcohol, and marijuana, and use one year earlier. Results showed an overall tendency for students to recall less use of uncontrolled substances than had been previously reported. For the one controlled substance included in the questionnaire, marijuana, current nonusers tended to recall less use than they had reported at the time, whereas current users tended to recall more use than had been reported. The present study found no evidence for a response-shift bias. It is suggested that the explicitly worded anchors on the response scales helped prevent such a bias. Finally, the results suggest that current use biases recall of past use to a substantial extent, and that this bias affects recall of alcohol use most severely.

Retrospective reports of behavior are used mainly under two sets of circumstances. First, retrospective reports may be used as a substitute when concurrent data are not available. For example, some-

times crucially relevant phenomena, for example, the drinking behavior of an alcoholic being admitted to treatment, have taken place before the beginning of a study or treatment program. Another instance when retrospective reports could be used as a substitute for concurrent data is in panel studies. If a subject is absent for an intermediate wave of data collection, a retrospective account collected at a subsequent wave could be inserted to replace the missing observation (Finney, 1981).

The second set of circumstances under which retrospective reports of behavior are used is when a response-shift bias may be operating (Howard, Millham, Slaten, & O'Donnell, 1981; Howard, Ralph, Gulanick, Maxwell, Nance, & Gerber, 1979). The response-shift bias affects pretest-posttest designs in which a treatment group receives an intervention intended to alter behavior. When a behavioral intervention alters not only the treatment subjects' behavior, but also their awareness of the behavior and their judgment about their own level of skill, the treatment subjects may undergo a change in their interpretation of the anchors of a response scale. This change is what is meant by response-shift bias. Howard et al. (1979) suggested that when a response-shift bias is operating, retrospective reports can be used to obtain a more nearly accurate assessment of pretest-posttest change.

Accuracy of Retrospective Reports

Although relatively little research has been done on the advisability of substituting retrospective re-

301

ports for concurrent reports, there is some evidence that retrospective reports can be adequate indicators of past behavior. Maisto, Sobell, Cooper, and Sobell (1982) compared two interview techniques for obtaining retrospective reports of drinking-related behavior from alcohol abusers. Examination of individual items included in their study reveals that more specific questions, such as those asking how many days the subject spent in jail for alcohol-related offenses in the last 30 days, tended to show a higher degree of agreement. In an investigation of agreement between recalled data and past reported data, Finney (1981) added a set of questions to the second wave of a panel study. These new questions asked for recall on several items included in the first wave of measurement four years earlier. For example, one question was, "Back in 1967, would you say your own health in general was excellent, good, fair, or poor?" Finney found that subjective questions asking for evaluations, such as the health question above, showed less agreement between first-wave response and recall than did more objective questions, such as whether or not the subject grew his or her own food. Taken together, the Finney (1981) and Maisto et al. (1982) studies suggest that moderate to high recall accuracy is possible when retrospective questions are relatively specific and objective.

Another study, however, found less encouraging results. Simpura and Poikolainen (1983) asked Finnish men to recall their alcohol use 18 years ago, and compared the recall with alcohol consumption reported in an interview occurring 18 years previously. They found that the recalled use tended to be considerably higher than the subjects' reported use at the first interview, and overall agreement between retrospective and concurrent reports was quite low. Of course, the original report and later recall in this study were separated by 18 years, which may be an unreasonably long time over which to expect subjects to remember. Simpura and Poikolainen considered the possibility that the drastic increase in per capita consumption of alcohol in Finland since the first interview may have affected the respondents' perceptions of their past habits. However, they found in regression analyses that current use did not predict recalled use. Although

no evidence for bias attributable to current use was found in this study, the possibility that current use influences recall of use should not be discarded. Simpura and Poikolainen's N was only 67, which allowed only a weak test of the hypothesis that current use affects recall of past use.

The Present Study

The present study addressed several interesting questions raised by this literature. The first question was: To what extent does recall of past substance use agree with past concurrent reports of substance use? In the case of alcohol, the Maisto et al. (1982) study leads to the expectation of a great deal of agreement, whereas Simpura and Poikolainen (1983) suggested that there may be relatively little agreement. The present study examined the correspondence between past reported use and recalled use for tobacco and marijuana as well as for alcohol.

In the present study, subjects took part in a smoking prevention program for high school students (Johnson, Hansen, Collins, & Graham, 1984). Thus, the second question addressed by the present study was: Will a smoking prevention program produce a response-shift bias? It is possible that an increased awareness of smoking behavior due to the intervention could cause a response-shift bias in reported tobacco use. Because some diffusion of treatment effects over alcohol and marijuana were found (Johnson et al., 1984), reported use of these substances was examined for response-shift bias as well.

A third question addressed here was: Does current substance use bias recall of past substance use? Simpura and Poikolainen (1983) raised this interesting possibility, and though they concluded that there were no biasing effects of current use in their study, the larger N of the present study allows a more powerful test of the hypothesis.

Method

Subjects were high school students in the Los Angeles area who were participants in the High School Anti-Smoking Project (HASP). As part of this project, students completed a questionnaire in

January 1981 and May 1982. The questionnaire contained items asking about tobacco, alcohol, and marijuana use. Three questions (1, 2, and 3) asked about use habits, and four questions (4, 5, 6, and 7) asked if subjects had ever tried a substance. The questions and response categories are shown in Table 1.

In the spring of 1983 the questionnaire was administered again, with the addition of a supplemental questionnaire. The supplemental questionnaire asked subjects to recall their tobacco, alcohol, and marijuana use habits one year earlier, that is, at the time of the May 1982 questionnaire, and 2½ years earlier, that is, at the time of the January 1981 questionnaire. Each substance use question and set of response categories was worded as similarly as possible to its earlier counterpart, so as to maximize the possibility of exact recall. The rewording of the questions also appears in Table 1. A total of 415 students who were present at the January 1981 data collection and the May 1982 data collection also completed the supplemental questionnaire. However, due to missing data on individual items, the N varied slightly among analyses.

Results

Overall Agreement Between
Past Reported Use and Recalled Use

Overall agreement between past reported use and recalled use was examined in three ways: (1) percent exact agreement between reported use and recalled use; (2) a measure of association (r) for the use habits items, and Cohen's kappa (1968) for the dichotomous "ever tried" items; and (3) t-tests of the mean differences between reported use and recalled use. These results are presented in Table 2 for the use habits items and Table 3 for the "ever tried" items. For both the use habits items and the "ever tried" items, marijuana showed the greatest degree of agreement, while both the alcohol use habits and "ever tried" items showed the lowest degree of agreement.

Differences between mean reported use and mean recalled use were significant in every case except for the marijuana use habits item involving use 2½

years ago. In every case, recalled use is less than was reported at the time, with one exception: Interestingly, subjects recalled more marijuana use than they had reported one year earlier.

Response-Shift Bias in
Reported Substance Use

In a pretest-posttest design, presumably there are no between-group differences at pretest in interpretation of the dependent measure response scales. If the treatment produces a response-shift bias, then at posttest the treatment subjects use the response scale differently from control subjects. When asked to recall pretest behavior, treatment subjects are now using a response scale different from the one on which they based their original pretest report. It follows from this that if a response-shift bias occurs, then the mean difference between pretest and recalled pretest should vary by condition. Accordingly, t-tests were used to test for differences between the treatment and control groups on the disparity between reported pretest use habits and later recalled use habits. No significant differences were found for tobacco, alcohol, or marijuana use habits (all $ps > .15$).

Bias in Recall
Due to Current Use

If both the reported past use and recall measures used in this study were perfectly valid and reliable, then the correlation between them would be 1.0. In the present study this correlation is less than perfect, as Table 2 shows, meaning that some of the variance in recall is left unaccounted for by reported past use. This leftover variance may not be all random error. If current use is a significant biasing factor in recalled use, then current use should account for some of this leftover variance; that is, current use should reliably predict the variance in recalled use that is left *after* the effects of reported past use have been partialled out.

The model of the present study is the following:

$$\text{Recalled Use} = \text{Past Use} + \text{Current Use}$$

Any variance in recalled use that is attributable to

Table 1
Questionnaire Items Used in the Present Study

Original Question	One Year Recall
1. Have you ever smoked even part of a cigarette? a. No, I have never tried it. b. Yes, I have tried it once or twice. c. Yes, I smoke less than a pack a week. d. Yes, I smoke about a pack a week. e. Yes, I smoke more than a pack a week but less than a pack a day. f. Yes, I smoke about a pack a day.	1. Think back to May 1982. At that time, had you ever smoked even part of a cigarette? a. No, at that time I had never tried it. b. Yes, at that time I had tried it once or twice. c. Yes, at that time I smoked less than a pack a week. d. Yes, at that time I smoked about a pack a week. e. Yes, at that time I smoked more than a pack a week but less than a pack a day. f. Yes, at that time I smoked about a pack a day.
2. How often do you drink alcoholic beverages? a. I never drink alcohol. b. Less than once a month. c. Once a month or more but less than once a week. d. About once a week. e. More than once a week but not every day. f. Almost every day.	2. Think back to May 1982. At that time, how often did you drink alcoholic beverages? a. At that time I never drank alcohol. b. At that time I drank alcohol less than once a month. c. At that time I drank alcohol once a month or more but less than once a week. d. At that time I drank alcohol about once a week. e. At that time I drank alcohol more than once a week but not every day. f. At that time I drank alcohol almost every day.
3. How often do you smoke marijuana? a. I never smoke marijuana. b. Less than once a month. c. Once a month or more but less than once a week. d. About once a week. e. More than once a week but not every day. f. Almost every day.	3. Think back to May 1982. At that time, how often did you smoke marijuana? a. At that time I never smoked marijuana. b. At that time I smoked marijuana less than once a month. c. At that time I smoked marijuana once a month or more but less than once a week. d. At that time I smoked marijuana about once a week. e. At that time I smoked marijuana more than once a week but not every day. f. At that time I smoked marijuana almost every day.
4. Have you ever tried drinking beer? a. Yes b. No	4. Think back to May 1982. At that time, had you ever tried drinking beer? a. Yes, at that time I had tried it. b. No, at that time I had never tried it.
5. Have you ever tried drinking wine? a. Yes b. No	5. Think back to May 1982. At that time, had you ever tried drinking wine? a. Yes, at that time I had tried it. b. No, at that time I had never tried it.
6. Have you ever tried drinking other alcoholic drinks? a. Yes b. No	6. Think back to May 1982. At that time, had you ever tried drinking any other alcoholic drinks? a. Yes, at that time I had tried other alcoholic drinks. b. No, at that time I had never tried any other alcoholic drinks.
7. Have you ever tried smoking marijuana? a. Yes b. No	7. Think back to May 1982. At that time, had you ever tried smoking marijuana? a. Yes, at that time I had tried smoking marijuana. b. No, at that time I had never tried smoking marijuana.

Table 2
Agreement Between Past Reported Substance Use Habits
and Recalled Use Habits

	Percent Agreement	r [a]	Mean Reported	Mean Recalled	t
Recalling One Year Ago					
Tobacco	63.6	.815	1.051	0.859	4.93 [c]
Alcohol	53.5	.702	1.384	1.281	2.19 [b]
Marijuana	74.1	.760	0.746	0.844	-2.14 [b]
Recalling 2 1/2 Years Ago					
Tobacco	61.0	.618	0.928	0.713	4.40 [c]
Alcohol	47.7	.532	1.158	0.992	3.21 [c]
Marijuana	69.3	.690	0.750	0.713	0.71

[a] All correlations are significant at p<.001
[b] p<.05
[c] p<.01

past use, including the part of past use that is correlated with current use, is considered legitimate. Any additional independent contribution of current use is considered bias.

Accordingly, a series of hierarchical regressions predicting recalled use was done, with past use always entered into the model before current use. This order of inclusion allowed all of past use, including the part that overlaps with current use, to account for variance in recalled use.

The results appear in Table 4. In these analyses past use always accounts for more of the variance than current use does, as might be expected. However, current use explains a significant amount of additional variance in every case. For one-year recall, the contribution of current alcohol use is strongest, both in absolute and in relative terms, when compared to the contribution of past use. For 2½-year recall, the contribution of current tobacco use is strongest in absolute terms, but current alcohol use accounts for relatively more of the variance in recalled alcohol use when compared to the contribution of past use. For all three substances, the variance attributable to current use increases from the one-year interval to the 2½-year interval, both in terms of absolute R^2 and in terms of the relative contribution to overall R^2.

The analyses shown in Table 4 include all stu-

Table 3
Agreement Between Past Reported Substance Trying
and Recalled Substance Trying

Substance	Percent Agreement	κ *	Percent Reported	Percent Recalled	t *
Recalling One Year Ago					
Tobacco	77.4	.553	66.7	48.7	8.05
Beer	85.6	.539	85.1	76.6	4.57
Wine	87.6	.559	87.8	78.9	5.26
Liquor	82.3	.529	79.2	71.3	3.80
Marijuana	89.1	.782	52.5	46.0	3.99
Recalling 2 1/2 Years Ago					
Tobacco	75.6	.522	61.5	45.9	6.59
Beer	79.4	.460	81.4	68.8	5.74
Wine	80.7	.439	82.4	74.0	3.92
Liquor	73.6	.403	73.1	62.3	4.16
Marijuana	82.7	.640	42.6	36.8	2.79

[a] p<.01 for all κ's and all t's

Table 4
Contributions (R^2 x 100) of Past Use and Present Use
to the Prediction of Recalled Use
When Past Use is Entered First*

Substance	R^2 Due to Past Use	R^2 Due to Current Use
Recall of One Year Ago		
Tobacco	62.9	8.5
Alcohol	48.7	14.2
Marijuana	57.1	13.1
Recall of 2 1/2 Years Ago		
Tobacco	36.0	18.8
Alcohol	28.4	18.0
Marijuana	46.6	15.1

*All R^2s are significant at p<.001

dents. However, for many of these students substance use did not change from the past to the present, as measured concurrently at each wave. For these students, the correlation between past use and current use is perfect, so after past use is entered into a regression equation predicting recalled use, current use adds nothing. For students whose current use is different from past use, the contribution of current use to recalled use will probably be larger than that shown in Table 4. Therefore, the regression analyses shown in Table 4 were redone including only students whose use habits changed. These results are shown in Table 5.

For this subset of students, the contribution of current use is larger than it is for the group as a whole, ranging from 19% to 25% of the variance at the one-year interval and from 23% to 28% of the variance at the 2½-year interval. The R^2 attributable to past use is lower overall for the students whose use habits changed. Based on these students, only the one-year recall of tobacco use appears to be relatively uncontaminated by current use. All other recall measures are considerably biased, particularly the 2½-year recall of alcohol use habits, in which the contribution of current use was twice that of past use.

Discussion

Overall Accuracy of Retrospective Reports

The present study showed that there was an over-

all tendency for students to recall less use than originally reported for the two uncontrolled substances, tobacco and alcohol. One possible explanation for this is that young students prefer to think of their substance use as a relatively recent development rather than a deeply ingrained habit. Thus, they may tend to deny or minimize past levels of use. Another possible explanation is that as time elapses and memory fades, students may fail to remember substance use occasions, and simply think their use was less because it seems so in retrospect.

For marijuana, however, students recalled more use than was reported one year ago, with no difference between recalled use and reported use 2½ years earlier. One possible explanation for this lies in an important difference between alcohol and tobacco on the one hand and marijuana on the other. Because marijuana is a controlled substance, it may be that past levels of use were underreported due to fears that the experimenters would single out students and report them to the authorities. By the last of four waves of measurement, which is when the recall instrument was administered, the students can be more confident that this will not happen. In addition, it is even less likely that reports of use a year or two ago will cause trouble for them. Thus, the students who were actually using marijuana one year or 2½ years ago may be likely to recall using more marijuana than they reported at the time. Because the past reported marijuana use is likely to be incorrect, this hypothesis cannot be tested by dividing the students into past users and past non-

Table 5
Contributions (R^2 x 100) of Past Use and Present Use to
the Prediction of Recalled Use, Including Only
Subjects Whose Use Habits Changed*

Substance	R^2 Due to Past Use	R^2 Due to Current Use	N
Recall of One Year Ago			
Tobacco	49.9	19.0	93
Alcohol	24.0	23.8	167
Marijuana	23.7	25.0	112
Recall of 2 1/2 Years Ago			
Tobacco	26.9	28.1	136
Alcohol	9.5	23.0	204
Marijuana	25.3	25.1	145

*All R^2s are significant at p<.01

users; however, this can be approximated by dividing the group into current users and current nonusers.

This division was done and an analysis of variance was performed with one between-subjects factor, current use, and two within-subjects factors, type of measure (report or recall) and time (one year ago or 2½ years ago). The analysis found a significant current use by type of measure interaction ($F[1,378] = 12.21$, $p < .001$) consistent with the hypothesis that marijuana users tend to recall *more* use than originally reported, whereas nonusers do not show this tendency. In fact, there was no difference between reported use and recalled use of one year ago for nonusers; however, for 2½-year recall, current nonusers showed the same tendency to recall less use than previously reported that was in evidence for tobacco and alcohol.

Thus for a controlled substance, marijuana, current nonusers recalled less use than previously reported, whereas current users recalled *more* use than previously reported. These currently-using students may still be recalling less marijuana use than they actually experienced, but more than they felt comfortable reporting earlier. Another possibility is that marijuana users may exaggerate past use because they feel that marijuana use is a desirable or high-status activity. However, this does not explain why such tendencies would exist for marijuana and not for alcohol or tobacco use.

Response-Shift Bias

The present study found no evidence for a response-shift bias in these data. A likely reason for the null results is the type of response scale anchors used. The present study used very explicit anchors, expressing substance use behavior in terms of daily, weekly, and monthly frequency. The work of Howard and colleagues (Howard et al., 1979; Howard et al., 1981) generally used less explicit anchors requiring subjective judgments, such as "average" and "good." Although the present study did not provide a direct comparison of explicit versus nonexplicit response-scale anchors, it is predicted that such a study would find a much smaller response-shift bias effect if explicit anchors were used. This conjecture is supported by Maisto et al. (1982) and Finney (1981), who found better recall accuracy using explicit questions.

Current Substance Use Bias of Recalled Substance Use

Finally, the data showed some interesting results bearing on the issue of possible bias of recalled substance use by current substance use. When all students were considered, the bias was significant, but relatively small, ranging from about 12% to 22.6% of the total R^2 at the one-year interval, and from 25% to 29% of the total R^2 at the 2½-year interval. However, calculations based only on stu-

dents whose use changed from wave to wave showed that 28% to 51% of the total R^2 was attributable to current use for one-year recall, and from 50% to 71% of the total R^2 was attributable to current use for 2½-year recall. These data also suggest that the current use bias operates in alcohol use recall to a particularly high degree. These results make the somewhat ironic suggestion that retrospective reports of use habits are most valid when use habits have not undergone change since the time period being recalled.

Implications

The results of the present study have methodological implications in three areas. First, these results suggest that experimenters should use caution in basing use rates of uncontrolled substances on retrospective self-reports. These rates are likely to be lower than would be obtained with concurrent reports. However, comparisons made between groups are likely to be valid, as long as the absolute rates are not interpreted. For controlled substances, retrospective reports may actually be more valid than a concurrent report, if the subject has any reason to fear legal ramifications from his or her answers. Perhaps some future study will speak to this issue more directly. For instance, it is possible that "bogus pipeline" techniques may be useful in encouraging more accurate concurrent reporting of marijuana use as has been suggested for tobacco self-reports (Evans, Hansen, & Mittelmark, 1977; Luepker, Pechacek, Murray, Johnson, Hurd, & Jacobs, 1981).

Second, the present study suggests that the use of objective, explicit anchors on response scales may help reduce or eliminate response-shift bias. Of course, such anchors do not make sense under all circumstances. Perhaps a future study will provide a direct comparison of response-shift bias produced with explicit versus nonexplicit anchors.

Finally, the results showed that retrospective accounts of past use habits are likely to be biased to a substantial degree by current use habits, especially if use habits have undergone change. The recall of alcohol use habits seems to be especially susceptible to this bias. The present results imply that in an adult population whose use habits have stabilized, retrospective reports may be quite valid. However, in a school-age or even college-age population whose use habits are likely to be changing rapidly, retrospective reports may be relatively invalid. Since alcohol use habits probably vary over time in any population more than most other substance use habits, alcohol use recall may be of extremely limited research utility.

References

Cohen, J. (1968). Nominal scale agreement with provision for disagreement or partial credit. *Psychological Bulletin, 70,* 213–220.

Evans, R. I., Hansen, W. B., & Mittelmark, M. B. (1977). Increasing the validity of self-reports of behavior in a smoking-in-children investigation. *Journal of Applied Psychology, 62,* 521–523.

Finney, H. C. (1981). Improving the reliability of retrospective survey measures: Results of a longitudinal field survey. *Evaluation Review, 5,* 207–229.

Howard, G. S., Millham, J., Slaten, S., & O'Donnell, L. (1981). Influence of subject response style effects on retrospective measures. *Applied Psychological Measurement, 5,* 89–100.

Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement, 3,* 1–23.

Johnson, C. A., Hansen, W. B., Collins, L. M., & Graham, J. W. (1984). *High school smoking prevention: Results of a three-year longitudinal study.* Manuscript submitted for publication.

Luepker, R. V., Pechacek, T. F., Murray, D. M., Johnson, C. A., Hurd, F., & Jacobs, D. R. (1981). Saliva thiocyanate: A chemical indicator of smoking in adolescents. *American Journal of Public Health, 71,* 1320–1324.

Maisto, S. A., Sobell, L. C., Cooper, A. M., & Sobell, M. B. (1982). Comparison of two techniques to obtain retrospective reports of drinking behavior from alcohol abusers. *Addictive Behaviors, 7,* 33–38.

Simpura, J., & Poikolainen, K. (1983). Accuracy of retrospective measurement of individual alcohol consumption in men: A reinterview after 18 years. *Journal of Studies on Alcohol, 12,* 911–916.

Acknowledgments

Author's Address

Send requests for reprints or further information to Linda M. Collins, Health Behavior Research Institute, University of Southern California Health Sciences Campus, 1985 Zonal Ave., Los Angeles CA 90033 U.S.A.