

5-1-1988

The Reliability of Alcohol Abusers' Self-Reports Of Drinking and Life Events That Occurred In the Distant Past

Linda C. Sobell

Nova Southeastern University, sobelll@nova.edu

Mark B. Sobell

Nova Southeastern University, sobellm@nova.edu

Diane M. Riley

University of Toronto

Reinhard Schuller

D. Sigfrido Pavan

University of Toronto

See next page for additional authors

Follow this and additional works at: http://nsuworks.nova.edu/cps_facarticles

 Part of the [Psychology Commons](#)

NSUWorks Citation

Sobell, L. C., Sobell, M. B., Riley, D. M., Schuller, R., Pavan, D. S., Cancilla, A., Klajner, F., Leo, G. I. (1988). The Reliability of Alcohol Abusers' Self-Reports Of Drinking and Life Events That Occurred In the Distant Past. *Journal of Studies on Alcohol*, 49(3), 225-232. Available at: http://nsuworks.nova.edu/cps_facarticles/696

This Article is brought to you for free and open access by the College of Psychology at NSUWorks. It has been accepted for inclusion in College of Psychology: Faculty Articles by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.

Authors

Linda C. Sobell, *Nova Southeastern University*, sobelll@nova.edu

Mark B. Sobell, *Nova Southeastern University*, sobellm@nova.edu

Diane M. Riley, *University of Toronto*

Reinhard Schuller

D. Sigfrido Pavan, *University of Toronto*

Anthony Cancilla

Felix Klajner, *University of Toronto*

Gloria I. Leo

The Reliability of Alcohol Abusers' Self-Reports of Drinking and Life Events that Occurred in the Distant Past*

LINDA C. SOBELL,† MARK B. SOBELL,† DIANE M. RILEY,† REINHARD SCHULLER, D. SIGFRIDO PAVAN,† ANTHONY CANCELLA, FELIX KLAJNER† AND GLORIA I. LEO

Addiction Research Foundation, Clinical Institute, 33 Russell Street, Toronto, Ontario M5S 2S1, Canada

ABSTRACT. This study investigated the test-retest reliability of 69 alcohol abusers' current reports about their past (approximately 8 years prior to interview) drinking behavior and life events. Drinking behavior was assessed by the Lifetime Drinking History (LDH) questionnaire and life events were assessed using the Recent Life Changes Questionnaire (RLCQ). Reliability coefficients for LDH variables were generally moderate to high ($r = .52$ to $.81$). Using empirical criteria, the diagnostic power of the two LDH interviews to classify correctly subjects as either having had or not having had a drinking problem was quite high. The reliability coefficient for the RLCQ was $r = .85$ and 91.7% of the identified events were reported in both interviews. Similarly high test-retest reliabilities and individual

event agreement rates were obtained for the six homogeneous subscales of the RLCQ. Subjects were also asked why they had given inconsistent answers to life events questions in the two interviews. Inconsistencies often resulted from errors in the temporal placement of events or from misunderstanding items, rather than from failure to recall an event; this suggests that some sources of error in recalling life events can be reduced. It is concluded that alcohol abusers' reports of drinking and life events occurring many years prior to the date of interview are generally reliable. This finding is consistent with previous studies showing high test-retest reliabilities for reports of recent drinking and related events. (*J. Stud. Alcohol* 49: 225-232, 1988).

WHILE NUMEROUS INVESTIGATORS (Polich, 1982; Sobell and Sobell, 1982, 1986) have found that alcohol abusers' self-reports of recent (i.e., within a year prior to the interview) drinking and related behaviors are generally reliable and valid when obtained under appropriate conditions, very little is known about the reliability of reports regarding events or behaviors that occurred in the distant past. What little research exists, however, suggests that for alcohol abusers certain distant information can be obtained fairly reliably (e.g., lifetime drinking history: Skinner and Sheu, 1982; questionnaires assessing fears and the severity of alcohol dependence for periods of fewer than 10 years past: Stockwell et al., 1984). In the alcohol field, assessment of distant events has gained importance as research has

begun focusing on the stability and patterning of long-term treatment outcomes (e.g., Taylor et al., 1985) and on the relationship of various events to either long-term natural recovery (e.g., Tuchfeld, 1981) or lifetime drinking history (e.g., Stockwell et al., 1984). Such research requires subjects to recall drinking and related events that occurred several years prior to the interview date, and the reliability of such reports was the topic of the present study.

The instrument used to assess drinking behavior in the present study was the Lifetime Drinking History (LDH) questionnaire (Skinner and Sheu, 1982). The LDH requires that subjects recall their lifetime drinking in discrete phases (including the problem period) involving major changes in their average drinking pattern. Although the LDH has been shown to be reliable in assessing lifetime drinking history, it is unknown whether this instrument can reliably be used to recall only selected phases (e.g., problem period) of a person's drinking history.

For the past 2 decades, several life events questionnaires have been used in clinical research studies; yet, despite their popularity, few studies have examined their psychometric characteristics, and very little research has been conducted examining the reliability of life events questionnaires covering distant time periods (i.e., more than 3 years from the interview date). Moreover, the

Received: 1 December 1986. Revision: 7 June 1987.

* Data presented in this article were presented, in part, at the annual meetings of the Association for Advancement of Behavior Therapy, Houston, Texas, November 1985, and the Canadian Psychological Association, Toronto, Canada, June 1986. The views expressed in this study are those of the authors and do not necessarily reflect those of the Addiction Research Foundation.

† Dr. L. Sobell and Dr. M. Sobell are also with the Departments of Psychology and Behavioural Science, University of Toronto. Dr. Riley and Dr. Klajner are also with the Department of Psychology, University of Toronto. Mr. Pavan is with the Department of Engineering, University of Toronto.

findings of studies investigating the reliability and validity of life events scales have generally not been encouraging (i.e., low to moderate reliability and validity) (Neugebauer, 1984; Thoits, 1983; Zimmerman, 1983). The life events scale used in the present study was the Recent Life Changes Questionnaire (RLCQ) (Rahe, 1975), an expanded version of the Social Readjustment Rating Questionnaire (SRRQ) (Holmes and Rahe, 1967; Rahe et al., 1964) and one of the more frequently used life events questionnaires.

While some investigators (Brown and Harris, 1978; Loftus and Marburger, 1983; Monroe, 1982; Paykel, 1983) have used various techniques to reduce life events recall errors due to temporal misclassifications (e.g., using calendars and diaries; having subjects attend to holidays and personally significant dates and events in the target recall period), surprisingly little research has been undertaken to determine reasons for unreliable answers. One study (Schless and Mendels, 1978) that addressed this question found that the most frequent reasons for disagreement between subject and informant reports of life events (as evaluated by a second informant) were (1) forgetting, (2) not thinking of the event, (3) forgetting that the event occurred in the time period under study, or (4) thinking that the event was unimportant. In light of these findings, perhaps one way to reduce life events recall error is to understand why subjects give unreliable answers (e.g., if subjective judgments of the importance of events is a major source of error, then providing subjects with guidelines for making such judgments could reduce the occurrence of these inconsistencies). To this end, Neugebauer (1984) has suggested that "research on checklist reliability should not provide merely overall estimates of measurement error, but information on the specific sources of that error" (*p.* 105).

The present study had two objectives. First, to examine the test-retest reliability of alcohol abusers' self-reports of drinking behavior and life events for a 1-year interval that occurred several years prior to the interview date. Second, to examine subjects' reported reasons for why they gave inconsistent answers (i.e., unreliable) to life events questions.

Method

Subjects

Sixty-nine patients (62 males, 7 females), recruited from inpatient and outpatient units at the Addiction Research Foundation in Toronto, participated as subjects in the present study. When recruited, subjects were informed that they would be participating in a pilot study to determine what kinds of questionnaires should be used in a future research study and that they would be interviewed on two different occasions and paid \$3 at the end of each interview (\$6 total).

To be eligible for the study, subjects had to: (1) sign an informed consent; (2) report alcohol as their major substance of abuse; (3) be at least 26 years of age (so they would be recalling events for a period when they were of legal drinking age); and (4) have no detectable blood alcohol level (BAL) when interviewed as determined by a breath test.

The sociodemographic characteristics of the 69 subjects who completed both interviews were as follows: (1) mean (\pm SD) 41.6 \pm 9.5 years of age; (2) 90% men; (3) 94% White; (4) 11.3 \pm 2.8 years of education; (5) 30% married; (6) 71% currently employed; and (7) 78% blue-collar work as their usual occupation. General drinking pattern data for these subjects were: (1) drinking problem history 10.9 \pm 9.4 years; (2) public drunk arrests 6.0 \pm 36.1 (median = 0; range = 0-300); (3) drunk driving arrests 1.1 \pm 1.5; (4) alcohol-related hospitalizations 2.4 \pm 6.9; (5) 68% reported blackouts; (6) 20% reported delirium tremens; and (7) 29% reported hallucinations. Nine additional subjects (all men) participated in the first session, but did not appear for their second interview. Those subjects who completed both interviews (n = 69), as compared with those who completed only the first interview (n = 9), differed on only one of the above variables (hallucinations; p < .02; 67% of noncompleters reported hallucinations).

Procedures

Subjects participated individually in two sessions scheduled to occur 2-3 weeks apart. The mean (\pm SD) test-retest interval was 16.8 \pm 3.4 days (range = 13-29 days; median = 15 days). Subjects were interviewed by one of four interviewers, with different interviewers conducting the first and second interviews. The interviewer who conducted a subject's second interview was blind to the subject's first interview answers.

Session 1. When potential subjects appeared for their first session, they read and signed an informed consent form and were then screened according to the eligibility criteria described earlier. After subjects answered a set of questions concerning their background and general drinking history, they were told that the period of time over which they were to recall their drinking behavior and life events was from January 1, 1976 through December 31, 1976. The year 1976 was arbitrarily chosen as the target year in order to reflect events occurring in the somewhat distant past. Prior to the first interview, subjects were not aware of the time period over which they would be asked to recall events.

Before completing the two questionnaires (LDH and RLCQ), subjects were told that, since they were being asked to recall events that had occurred 7½ to 8 years earlier, two memory aids would be used to assist their recall. The first involved viewing magazine covers depicting four major news events of 1976: Montreal Summer Olympics; death of Mao Tse-Tung; U.S.A. Bicentennial and President Jimmy Carter's election. The subjects were told that "the following pictures show major Canadian, world and sporting events that occurred in 1976. They are intended to help you better recall what you were doing and what was happening in 1976." The second memory aid was a "Personal Recall Calendar," which subjects completed prior to their interview and which asked

them for the following information about their life circumstances in 1976: (1) age at the beginning and end of the year; (2) living arrangements—with whom and where; (3) work—employer and job title; (4) if married—which wedding anniversary was celebrated; (5) ages of all children at the beginning and end of the year; and (6) number of years, if any, that drinking had been a problem up to and through 1976.

At the end of the first session, subjects were paid \$3 and the second interview was scheduled. Although informed that the second interview would consist of information gathering similar to the first session, subjects were not told they would be asked to report the same information.

Session 2. Subjects were asked to sign a second informed consent form indicating that they would be asked to answer the same questions as in the first interview. The consent form included the statement: "Repeating questionnaires on the same person is a standard part of assessing their usefulness." All subjects who appeared for the second session completed that session. One subject was not breath tested at the second interview, but he resided in an inpatient unit that prohibited drinking, and he told the interviewer the next day that he had not consumed alcohol on the day of the interview.

The procedures were identical to those of the first session, with two exceptions: the background and screening questions were not readministered, and, for 35 of the 69 subjects, further information was systematically gathered for all life events that were answered differently in the second interview. This latter procedure was initiated after it was observed that some subjects during the second interview spontaneously explained to the interviewers that they had reported a particular life event in only one of the interviews because they had, for example, made a mistake in dating the event. Since it is important to distinguish answers that are unreliable because of dating errors (i.e., the event occurred but was simply misplaced in time) from those that are unreliable because of forgetting, reasons for inconsistent answers were probed for the final 35 subjects. For these subjects, at the end of their second session, the interviewer opened a sealed envelope containing the subject's first interview answers and tabulated any inconsistent responses in the two interviews. Subjects were then asked to comment on each inconsistency, and, if possible, to provide a reason for it. The interviewer recorded the subjects' comments. The subjects were paid \$3 and asked not to discuss the nature of the study with other potential subjects until the study had been completed.

Instruments

Lifetime Drinking History. The LDH was administered to subjects as a structured interview (Skinner and Sheu, 1982). Although the target recall period covered only one year, data were obtained in the same way as if lifetime drinking history information were being sought. For each major phase of drinking in 1976, quantity (drinks/day) and frequency (days/month) data were obtained for the subject's average and maximum pattern. Although the original LDH did not collect frequency data for a subject's maximum pattern, the author of the scale, Dr. Skinner, suggested that we evaluate this variable. Subjects were asked to report their alcohol con-

sumption using a standard drink formula (1 standard drink = 12 oz of 5% beer, 1½ oz of 80% distilled spirits, 5 oz of 12% wine, or 3 oz of 20% wine; all contain approximately 13.6 g of absolute alcohol). Reports of morning drinking during each phase were also obtained.

Recent Life Changes Questionnaire. The full RLCQ consists of 76 items requiring "yes" or "no" answers. Some of the life events, however, have multiple parts (e.g., death of a: (a) child, (b) brother or sister, (c) parent, (d) other close family member); when each multiple-part item is collapsed into a single item (e.g., death of a close relative), a 55-item RLCQ (collapsed scale) is obtained. Also, for each life event checked "yes," subjects were asked to evaluate whether the event was a result of their drinking (yes, no) and what impact the event had on their life (positive, negative, no impact). Subjects were allowed to use two impact codes if they felt that was the correct answer (e.g., both a positive and negative impact on their life). Initially, it was planned to have subjects self-administer the RLCQ, but some subjects had difficulty completing or understanding the checklist, forgot their glasses, or asked the interviewer to help them complete the form. Thus, for 27 of the subjects, the RLCQ was administered by the interviewer. There is no evidence of any difference in reliability between these two methods of administration (Zimmerman, 1983).

Results

RLCQ reliability

Pearson correlation coefficients for test-retest reliability are shown in Table 1 for the full scale RLCQ score, for the collapsed scale RLCQ score (any response of "yes" to a multiple-part event resulted in the event being scored as having occurred) and for the six subscale scores making up the full and collapsed scales. Before computing the correlations, square root transformations were performed to reduce skewness in the original data (Cohen and Cohen, 1975).

The reliability estimates for both the full and collapsed scale RLCQ scores were .85 and .84, respectively, with subscale reliabilities somewhat lower (ranging from .60 to .76). Some of the subscales would be expected to have lower reliability coefficients than others since reliability is affected by scale length. Therefore, to increase the comparability among subscales, subscale scores were adjusted by the Spearman-Brown formula (Cronbach, 1970) to reflect a scale length of 18 items (i.e., the greatest number of items in any one subscale). The adjusted reliability estimates for the subscales also appear in Table 1, with values ranging from .68 to .85 (full) and from .76 to .88 (collapsed). Overall, the reliability coefficients for the collapsed subscales were slightly higher than their corresponding reliabilities for the full subscale scores.

Rate of concordance (percentage of agreement) provides direct information on individual event agreement

TABLE 1. Test-retest reliability correlations and percent of item agreement for the RLCQ scale and subscale scores^a

Variable	No. of items	Reliability estimate	Adjusted reliability estimate ^b	% Total agreement (item concordance)		
				For items comprising the scale	Drinking impact code ^c	Life impact code ^c
RLCQ full scale ^d	76 (55)	.85 (.84)	-	91.7 (91.0)	82.8	71.7
RLCQ subscales						
Health	6 (5)	.63 (.63)	.84 (.86)	82.6 (82.0)	73.8	55.0
Work	16 (8)	.66 (.63)	.68 (.79)	91.1 (95.3)	93.6	64.9
Home-family	13 (8)	.60 (.59)	.68 (.76)	93.1 (90.0)	81.0	73.7
Marriage	17 (12)	.73 (.73)	.74 (.81)	97.8 (96.9)	83.3	62.5
Personal-social	18 (18)	.76 (.76)	.76 (.76)	88.9 (88.9)	80.4	78.2
Financial	6 (4)	.65 (.62)	.85 (.88)	90.6 (87.7)	82.7	82.7

^a $p < .0001$ for all correlations.

^b Reliability adjusted for a scale length of 18 items by Spearman-Brown formula.

^c For items answered "yes" in both interviews.

^d Collapsed RLCQ scale scores are in parentheses.

between the two interviews. Table 1 shows the percentage of life events reported identically in both interviews. Concordance between specific events was 91.7% for the full scale and 91% for the collapsed scale. Comparably high individual event agreement was evident among the subscales for both the full and collapsed scales. The health subscale had the lowest overall concordance, perhaps reflecting a greater number of items requiring subjective interpretations. The high rates of concordance indicate that the correlation coefficients primarily reflected item agreement between interviews rather than simply systematic ordinal relationships.

The mean (\pm SD) numbers of life events reported by subjects in the first interview were 9.5 ± 6.5 (range: 0-27) for the full and 8.9 ± 6.3 (range: 0-28) for the collapsed scales. This compares favorably with the rate of reported events (mean=9.7) found in another study that used a similar population and a similar checklist (Skinner and Lei, 1980).

Table 1 also presents the percentage of concordant answers (agreement) for the two impact codes for items answered "yes" in both interviews. If a subject gave two impact code answers to an item (e.g., both positive and negative life impact), he or she had to have done that in both interviews for the answer to be scored as concordant. The overall percentage of concordant answers was somewhat higher for drinking impact codes (82.8%) than for general life impact codes (71.7%). The subscale concordance rates for the two impact codes similarly were higher for the drinking impact evaluations than for the general life impact evaluations. These findings suggest that although subjects may report the occurrence of life events fairly reliably and consistently, their subjective evaluations of the impact of the events are reported less consistently, especially for general life impact evaluations.

Sources of life event unreliability

Two trained raters independently coded subjects' reasons for inconsistent answers on the full scale RLCQ into five categories that had been defined by two of the authors after they had examined the subjects' original answers: (1) dating the event incorrectly; (2) forgetting or not recalling the event; (3) don't know (only response given); (4) misunderstood the question, mistakenly categorized an event, or re-evaluated an event's importance; and (5) other—misread question, not sure if event occurred, error, thought the same answer was given twice, reported the event but as part of another event (this latter example usually occurred for multiple part questions).

A Kappa coefficient was computed to measure interrater reliability between raters' categorizations of subjects' reasons for giving inconsistent (unreliable) answers in the first and second interviews (Cohen, 1960). The Kappa coefficient measures the extent of interrater agreement beyond that which would be expected by chance, and ranges from 1.0 (perfect agreement) to 0 (no agreement). With $\kappa = .896 \pm .024$ ($p < .001$), it is estimated that the chances are 99% that the population value of κ falls between .83 and .96.

When raters' codings were in agreement (92% of all reasons for inconsistent answers; 199 of 216), over half (54.3%) of those cases involved instances where subjects stated that on one occasion they had either dated the event incorrectly, misunderstood the question, mistakenly categorized their answer or re-evaluated the importance of the event (some RLCQ items required subjects to decide whether an event involved a "major" change). One quarter (25.1%) of the reasons for inconsistent answers involved subjects indicating that

TABLE 2. Means (\pm SDs) of test-retest drinking behavior data and correlation coefficients for subjects' reports of their average drinking pattern in 1976

Variable ^a	Session 1	Session 2	<i>r</i> ^b
	M \pm SD (Range)	M \pm SD (Range)	
Days any alcohol consumed (frequency:F)	213.1 \pm 118.3 (0-360)	208.7 \pm 120.7 (0-360)	.81 (.54) ^c
Total no. drinks (quantity:Q)	19.8 \pm 19.1 (0-84)	18.0 \pm 18.1 (0-120)	.66 (.69) ^c
Drinks/drinking day (Q-F)	10.8 \pm 7.7 (0-35)	10.4 \pm 7.9 (0-40)	.67 (.64) ^c
Greatest no. standard drinks on any single day	12.2 \pm 9.3 (0-48)	11.5 \pm 8.7 (0-40)	.65 (.67) ^c
Days 1-4 standard drinks	49.3 \pm 108.6 (0-360)	50.5 \pm 102.8 (0-360)	.53 ^{-d}
Days > 4 standard drinks	163.8 \pm 133.0 (0-360)	158.2 \pm 136.5 (0-360)	.52 ^{-d}
Days 1-6 standard drinks	78.6 \pm 119.3 (0-360)	77.0 \pm 122.9 (0-360)	.73 ^{-d}
Days > 6 standard drinks	134.5 \pm 134.2 (0-360)	131.7 \pm 135.8 (0-360)	.70 ^{-d}
Longest consecutive abstinent period	24.4 \pm 71.0 (0-360)	34.2 \pm 87.9 (0-360)	.88 (NA) ^e

^a During the 360 day reporting period.

^b All *p*'s < .0001.

^c Values in parentheses are correlation coefficients computed using data from subjects' reports of their maximum drinking pattern in 1976.

^d Since too few subjects reported maximum levels of drinking that would have been classified in the categories of 1-4 and 1-6 standard drinks, these correlation coefficients would have been either redundant or impossible to calculate. Thus, they were not computed.

^e NA = Not applicable.

they simply forgot to report the event on one occasion. Based on the current study, it appears that subjects can give specific reasons for the vast majority of their inconsistent answers.

LDH reliability

Table 2 presents descriptive statistics and product-moment test-retest reliability coefficients for several drinking-behavior variables derived from the LDH data. Before computing the correlations, a square-root transformation was conducted on the raw data to reduce skewness (Cohen and Cohen, 1975). The reliability coefficient for total days that any drinking occurred was .81. For purposes of analysis, drinking days were categorized as days abstinent, days 1-4 standard drinks (\leq 54 g absolute alcohol), days greater than 4 standard drinks, days 1-6 standard drinks (\leq 82 g absolute alcohol), and days more than 6 standard drinks were consumed. Two categories of limited drinking (i.e., 1-4 and 1-6 standard drinks) were included to evaluate how the reliability coefficients would differ as a function of how limited drinking was defined. As shown in Table 2, subjects appeared able to categorize their drinking more reliably when the categories were more broadly defined, as evidenced by the higher reliability coefficient for the category of days when 1-6 standard drinks were consumed ($r = .73$) compared with days when 1-4 were consumed ($r = .53$). Drinking behavior data for a subject's maximum consumption pattern were also examined for reliability (see Table 2).

Because the LDH was used differently in this study than in the original evaluation of the LDH (Skinner and Sheu, 1982), it was not possible to compare most

findings between studies. However, the reliability coefficient for a key variable in this study, average drinks per drinking day (quantity-frequency), was almost identical to that obtained by Skinner and Sheu, $r = .67$ and $r = .68$, respectively. Further, the other reliability coefficients in this study were either slightly higher or in the same range as those for the LDH variables examined by Skinner and Sheu.

Diagnostic utility of the two interviews

While the statistical analyses in the present study suggest that long-term retrospective reports of drinking behavior are generally reliable, an important question relating to the use of these data is whether one would form the same clinical impression regarding the seriousness of a subject's reported drinking problem from examining both sets of data. In order to increase the utility of clinical test data, researchers have started examining the predictive power of their test data. The following definitions are based on the assumption that the data from the first interview reflect the true state of affairs: (1) positive predictive power: the occurrence of a positive diagnosis at the first interview followed by the same diagnosis at the second interview (ratio of true-positives to all positives in the second interview), and (2) negative predictive power: the occurrence of a negative diagnosis (i.e., no alcohol problem) at the first interview followed by the same diagnosis at the second interview (ratio of true-negatives to all negatives in the second interview). For this study, as shown in Table 3, three different empirical definitions of an alcohol problem were used in the evaluation of the diagnostic and predictive power (see Baldessarini et al., 1983) of the data reported by the subjects in the two

TABLE 3. Diagnostic power of the two interviews to consistently classify subjects as having an alcohol problem, in percent

Variable	Alcohol problem diagnosis		Morning drinking
	Definition I ^a	Definition II ^b	
Overall diagnostic power:			
Ss correctly classified as having or not having an alcohol problem in both interviews	84	83	81
Positive predictive power:	91	86	72
T1 & T2 pos dx/T2 all pos dx ^c			
Negative predictive power:	57	72	84
T1 & T2 neg dx/T2 all neg dx ^d			

^a A monthly average consumption pattern of > 4 SDs or evidence in the maximum pattern of > 12 SDs on any day.

^b A monthly average consumption pattern of > 6 SDs or evidence in the maximum pattern of > 12 SDs on any day.

^c 1st interview and 2nd interview positive diagnosis/2nd interview all positive diagnoses.

^d 1st interview and 2nd interview negative diagnosis/2nd interview all negative diagnoses.

interviews. The diagnostic power of the two interviews to classify subjects consistently as either having or not having an alcohol problem in both interviews was found to be high ($\geq 81\%$) for all three definitions. In research with alcohol abusers it is more important to maximize the true positive results (i.e., correctly classify alcohol problems) as opposed to misclassifying those without alcohol problems. The positive predictive power of the interview data for the three alcohol problem definitions ranged from 72% to 91%.

Discussion

Using two assessment instruments, this study found that alcohol abusers' reports about drinking and life events that occurred in the distant past are generally reliable. This finding is important because several recent research studies in the alcohol field have required the recall of behaviors and events that occurred many years prior to the interview (e.g., long-term outcome evaluations, problem drinking history development and patterning, and evaluations of the natural course of alcohol problems).

With respect to drinking, this study found that the LDH questionnaire can reliably be used to gather subjects' recollections of their drinking during selected distant time periods in their drinking career. The drinking variables examined can be assessed with at least moderately high reliability, a conclusion similar to that reported by Skinner and Sheu (1982) when they had subjects use the LDH to summarize their entire drinking career. As indicated by the correlational analyses, subjects' reports of their drinking were not perfectly concordant between interviews. However, despite any differences that existed between the two interview data sets, the overall diagnostic accuracy and positive predictive power of the two interviews were very high. In summary, based on the reliability analyses and the

predictive and diagnostic power of the two interviews, the overall findings indicate that alcohol abusers' retrospective reports of their drinking behavior that occurred many years prior to the interview are generally reliable.

For alcohol abusers' reports of distant life events, the results of this study indicate that both the full and collapsed RLCQ scales, as well as their six subscales, had high test-retest reliability. Also, individual event concordance rates for the full and collapsed scales and six subscales were quite high. Comparisons of these latter figures with those from other studies is not possible since concordance rates for life events have seldom been reported in the literature. Although it was found that there was little difference between the reliability for the full and collapsed scale RLCQ scores, the full (76-item) RLCQ scale is recommended because it provides more information.

The test-retest concordance estimates for the two life events impact evaluations were somewhat lower than for the simple reporting of whether events occurred, but were still quite high, with drinking impact evaluations (82.8%) being slightly more reliable than life impact evaluations (71.7%). It is not possible to evaluate how these figures relate to the existing literature, since to our knowledge the concordance rates of life events impact codes have not been previously examined.

One unique aspect of this study involved questioning one-half of the subjects after their second interview about possible reasons for their inconsistent answers. Based on raters' evaluations of subjects' answers, it appears that more than one-half of the reasons for inconsistent answers could be classified into two of five categories: temporal misplacement and event misinterpretation or re-evaluation. Although one of the most commonly assumed sources of error in life events scale responses is forgetting (Funch and Marshall, 1984), subjects in the present study stated that

only 25% of the events not reported on one occasion had been forgotten. These findings are important, because they suggest that instances of poor reliability are more often the result of errors or relate to the wording of the life events items (e.g., requiring judgments as to whether events were "major") than the result of forgetting that the event occurred. This suggests that procedures might be developed in future research to enhance the reliability of responses to life events questionnaires.

For life events, three aspects of the present study may have contributed to the high reliabilities and high item concordance rates found. First, subjects in the present study were asked to recall events that had occurred several years prior to the interview. In this regard, some investigators have shown that event saliency is one of the most potent factors affecting recall consistency (Casey et al., 1967; Funch and Marshall, 1984). Thus, in the present study, the recall of events far removed in time might have acted as a filter to reduce recall of less salient events. Second, while it has been suggested that one source of error in life events recall involves overreporting or magnifying events to explain the onset of an illness (Rabin and Struening, 1976; Zimmerman, 1983), this type of error did not exist in the present study since the recall time frame was arbitrary and not related to significant events (e.g., illness onset). Third, the use of two memory aids (i.e., magazine covers and a personal recall calendar) may have helped subjects to recall reliably and to chronologize life events and reports of their past drinking behavior correctly. Similar aids have been used to help subjects recall recent drinking behavior (e.g., Sobell et al., 1979).

It should be noted that a Finnish study (Simpura and Poikolainen, 1983) has found poor reliability for recall of drinking using an 18-year test-retest interval. However, several major differences between that study and the present study make comparisons impossible. The Finnish study (1) did not use alcohol abusers as subjects, (2) used subjects from a different country than that used for the present study, (3) did not use memory aids to assist recall, and (4) used different data collection instruments between the two test-retest intervals (i.e., Time 1: intense and multiple data collection; Time 2: an aggregate one-time estimation formula).

In conclusion, the work described here complements existing reliability studies examining alcohol abusers' reports of recent drinking and related events. Although there is no reason to suspect that the present results are invalid, an evaluation of their validity would require comparison with an external criterion. While collateral confirmation is one possible criterion, because of the distant recall period and the social instability of many alcoholics, collateral informants may not be available or know about an alcohol abuser's distant

past. Also, while some life events (e.g., marriage, accidents) can be confirmed by official records, many life events, and certainly drinking behavior, cannot be so verified. Thus, when assessing alcohol abusers' reports of temporally removed behaviors and events, it is recommended that a convergent validity criterion be used to increase confidence in the data collected. A similar recommendation has been made when assessing the validity of reports of recent drinking behavior and events (Sobell and Sobell, 1986; Sobell et al., 1980). A convergent validity approach postulates that one can have confidence in the validity of data to the extent that those data are corroborated by a variety of alternative data sources. Finally since memory aids have not typically been employed in studies assessing life events or lifetime drinking, the present results may not generalize to studies not using such aids.

Acknowledgment

The authors thank Dr. Harvey Skinner for his comments on an earlier draft of this article.

References

- BALDESSARINI, R.J., FINKLESTEIN, S. AND ARANA, G.W. The predictive power of diagnostic tests and the effect of prevalence of illness. *Arch. gen. Psychiat.* **40**: 569-573, 1983.
- BROWN, G.W. AND HARRIS, T. *Social Origins of Depression: A Study of Psychiatric Disorder in Women*, New York: Free Press, 1978.
- CASEY, R.L., MASUDA, M. AND HOLMES, T.H. Quantitative study of recall of life events. *J. psychosom. Res.* **11**: 239-247, 1967.
- COHEN, J. A coefficient of agreement for nominal scales. *Educ. psychol. Meas.* **20**: 37-46, 1960.
- COHEN, J. AND COHEN, P. *Applied Multiple Regression Correlation Analysis for the Behavioral Sciences*, Hillsdale, N.J.: Lawrence Erlbaum Assocs., Inc., 1975.
- CRONBACH, J.L. *Essentials of Psychological Testing*, 3d Edition, New York: Harper & Row Pubs., Inc., 1970.
- FUNCH, D.P. AND MARSHALL, J.R. Measuring life stress: Factors affecting fall-off in the reporting of life events. *J. Hlth social Behav.* **25**: 453-464, 1984.
- HOLMES, T.H. AND RAHE, R.H. The social readjustment rating scale. *J. psychosom. Res.* **11**: 213-218, 1967.
- LOFTUS, E.F., AND MARBURGER, W. Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Mem. Cog.* **11**: 114-120, 1983.
- MONROE, S.M. Assessment of life events: Retrospective vs concurrent strategies. *Arch. gen. Psychiat.* **39**: 606-610, 1982.
- NEUGEBAUER, R. The reliability of life-event reports. In: DOHRENWEND, B.S. AND DOHRENWEND, B.P., (Eds.) *Stressful Life Events and Their Contexts*, New Brunswick, N.J.: Rutgers Univ. Press, 1984, pp. 85-107.
- PAYKEL, E.S. Methodological aspects of life events research. *J. psychosom. Res.* **27**: 341-352, 1983.
- POLICH, J.M. The validity of self-reports in alcoholism research. *Addic. Behav.* **7**: 123-132, 1982.
- RABIN, J.G. AND STRUENING, E.L. Life events, stress, and illness. *Science*, **194**: 1013-1020, 1976.
- RAHE, R.H. Epidemiological studies of life change and illness. *Int. J. Psychiat. Med.* **6**: 133-146, 1975.

- RAHE, R.H., MEYER, M., SMITH, M., KJAER, G. AND HOLMES T.H. Social stress and illness onset. *J. psychosom. Res.* **8**: 35-44, 1964.
- SCHLESS, A.P. AND MENDELS, J. The value of interviewing family and friends in assessing life stressors. *Arch. gen. Psychiat.* **35**: 565-567, 1978.
- SIMPURA, J. AND POIKOLAINEN, K. Accuracy of retrospective measurement of individual alcohol consumption in men: A reinterview after 18 years. *J. Stud. Alcohol* **44**: 911-917, 1983.
- SKINNER, H.A. AND LEI, H. The multidimensional assessment of stressful life events. *J. nerv. ment. Dis.* **168**: 535-541, 1980.
- SKINNER, H.A. AND SHEU, W.J. Reliability of alcohol use indices: The lifetime drinking history and the MAST. *J. Stud. Alcohol* **43**: 1157-1170, 1982.
- SOBELL, L.C., MAISTO, S.A., SOBELL, M.B. AND COOPER, A.M. Reliability of alcohol abusers' self-reports of drinking behavior. *Behav. Res. Ther.* **17**: 157-160, 1979.
- SOBELL, L.C. AND SOBELL, M.B. Alcoholism treatment outcome evaluation methodology. In: NATIONAL INSTITUTE ON ALCOHOL ABUSE AND ALCOHOLISM, Prevention, Intervention and Treatment: Concerns and Models. Alcohol and Health Research Monograph No. 3, DHHS Publication No. (ADM) 82-1192, Washington: Government Printing Office, 1982, pp. 293-321.
- SOBELL, L.C. AND SOBELL, M.B. Can we do without alcohol abusers' self-reports? *Behav. Therapist* **7**: 141-146, 1986.
- SOBELL, L.C., SOBELL, M.B. AND WARD, E. (Eds.) Evaluating Alcohol and Drug Abuse Treatment Effectiveness: Recent Advances, Elmsford, N.Y.: Pergamon Press, Inc., 1980.
- STOCKWELL, T., SMAL, P., HODGSON, R. AND CANTER, S. Alcohol dependence and phobic anxiety states. *II. A retrospective study.* *Brit. J. Psychiat.* **144**: 58-63, 1984.
- TAYLOR, C., BROWN, D., DUCKITT, A., EDWARDS, G., OPPENHEIMER, E. AND SHEEHAN, M. Patterns of outcome: Drinking histories over ten years among a group of alcoholics. *Brit. J. Addict.* **80**: 45-50, 1985.
- THOITS, P.A. Dimensions of life events that influence psychological distress: An evaluation and synthesis of the literature. In: KAPLAN, H.B. (Ed.) Psychosocial Stress: Trends in Theory and Research, Orlando, Fla.: Academic Press, 1983, pp. 33-103.
- TUCHFELD, B.S. Spontaneous remission in alcoholics: Empirical observations and theoretical implications. *J. Stud. Alcohol* **42**: 626-641, 1981.
- ZIMMERMAN, M. Methodological issues in the assessment of life events: A review of issues and research. *Clin. psychol. Rev.* **3**: 339-370, 1983.