Pergamon

Addictive Behaviors, Vol. 25, No. 1, pp. 1–11, 2000

0306-4603/00/$–see front matter

PII S0306-4603(99)00031-3

USING INTERACTIVE VOICE RESPONSE TECHNOLOGY AND TIMELINE FOLLOW-BACK METHODOLOGY IN STUDYING BINGE EATING AND DRINKING BEHAVIOR: DIFFERENT ANSWERS TO DIFFERENT FORMS OF THE SAME QUESTION?

ANNA M. BARDONE,* DEAN D. KRAHN,† BRIAN M. GOODMAN,‡ and JOHN S. SEARLES§

*University of Wisconsin-Madison; †University of Wisconsin-Madison and William S. Middleton Memorial Veterans Hospital; ‡Geriatric Research Education Clinical Center, William S. Middleton Memorial Veterans Hospital; and §Vermont Alcohol Research Center, University of Vermont

Abstract — As part of a study of the relationship of binge eating, alcohol use, mood, and stressors, we compared the results of two forms of reporting on binge eating and drinking behavior. Forty-three first-year college women participated in an interactive voice response (IVR) study for 12 weeks. Participants answered computer-administered questions daily via IVR technology on number of eating binges and number of alcoholic drinks consumed. After 12 weeks, participants completed a Timeline Follow-back (TLFB) interview retrospectively for number of binges and drinks in the past 12 weeks. Results of this distally retrospective methodology (commonly used in drinking research and applied here also to binge eating) were compared to the results of daily IVR reporting. There was convergence across measures for drinking behavior, but divergence between IVR and TLFB for binge eating reports. TLFB reports underrepresented actual binge eating frequency, which calls into question the validity of applying this methodology to the assessment of binge eating. © 2000 Elsevier Science Ltd

Key Words. Binge eating, Alcohol, Methodology, Interactive voice response, Timeline follow-back.

Clinicians and researchers of behavioral problems such as eating disorders and alcohol abuse generally require a report of behavior frequency in order to decide on a diagnosis or inclusion of individuals in a research program. For example, the binge frequency criterion of two or more times per week in the past 3 months is required for a diagnosis of bulimia nervosa or binge eating disorder (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition [*DSM-IV*]; American Psychiatric Association, 1994). In terms of alcohol use, physicians are advised to check for a threshold of seven or more drinks per week in their women patients as a warning sign for problem drinking (Adinoff, 1998), while binge drinking is defined using the frequency criterion of five or more drinks in a row (Johnston, O'Malley, & Bachman, 1994). It is important to note that these drinking guidelines are unrelated to *DSM-IV* criteria for alcohol abuse or dependence, but have been determined, in epidemiological research, to confer greater risk for later problems (Adinoff, 1998).

This research was supported in part by a grant from the Alcoholic Beverage Medical Research Foundation to Dean D. Krahn, and a National Science Foundation Fellowship to Anna M. Bardone. We thank the Vermont Alcohol Research Center, and in particular James C. Mundt and Dan Walter.

Requests for reprints should be sent to Anna M. Bardone, Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson Street, Madison, WI 53706; E-mail: abardone@students.wisc.edu

If criteria for binge eating (e.g., two or more binges/week) and problem alcohol use (e.g., 7 or more drinks/week) is important, then clinicians and researchers should understand how different means of assessing binge eating and drinking behavior affect the likelihood of a patient or study participant meeting the criteria. This assessment issue needs to be considered seriously because of the clinical, theoretical, and social significance of assigning an individual a mental disorder. Currently, the most common reporting methods for both binge eating and drinking require individuals to reflect over months and integrate information to arrive at an average rate. It is unclear how confident clinicians and researchers should be that patients and research participants accurately report binge eating and drinking behavior. When individuals respond to various methods of assessment, how well do their responses map onto their actual behavior?

In the context of a larger study designed to look at the temporal ordering of binge eating, drinking, mood, and stressors, we took the opportunity to address this methodological question. We used daily, confidential reporting of binge eating and drinking behaviors by a group of young women as the least retrospective and most accurate measure of daily behavior (a "gold standard") against which to compare a more distally retrospective assessment measure. In particular, daily reports on binge eating and drinking behaviors gathered using interactive voice response (IVR) technology over a 12-week period (84 days) were compared to retrospective reports gathered for that same 12-week period using Timeline Follow-back (TLFB) methodology. This design allowed us to test whether or not TLFB reports on drinking behavior and binge eating behavior agreed with IVR reports. Since we believe that more proximal reporting is more valid, a failure of agreement would suggest that TLFB is less valid.

Our objectives included:

1. To compare the distribution of binge eating and drinking behaviors across the two reporting methods for the same reporting period (including comparisons at levels considered potentially problematic, e.g., two or more binges/week, seven or more drinks/week,[1] five or more drinks in a row at least three times per 2-week period[2]).
2. To calculate the sensitivity and specificity of the TLFB methodology for identifying individuals who indicated potentially problematic levels of binge eating and drinking.

We hypothesize that there should be differences in reports of binge eating and drinking behaviors across the two methods used, with the retrospective report (i.e., TLFB) underestimating actual binge eating and drinking behaviors due to forgetting, inaccurate integration of information, and difference in subjective degree of anonymity.

[1]This frequency of drinking was chosen for reasons of prevalence as well as because it serves as a guideline for physicians (Adinoff, 1998). Participants who drank on average seven or more drinks per week represented the most heavily drinking one third of the sample. This rate of drinking was found in an even smaller fraction (about one fifth) of the female freshman college participants in a similar study (Kurth, 1994).

[2]This frequency of binge drinking was chosen because it represents the extreme 20% of the sample. Also, since binge drinking once a weekend may characterize many college drinkers, the frequency of three or more episodes of binge drinking in a 2-week period represents behavior that occurred in excess of once per weekend. Kurth (1994) found that only about one tenth of their college sample drank five or more drinks in a row at least three times in a 2-week period.

METHOD

Participants

During the summer before their college entry, 1,665 first-year female college students (ages 18–19) completed a general screening instrument assessing eating behaviors and attitudes, and drinking behaviors. Study participants were selected from the pool of 1,038 (62%) of the students who gave consent for future contact. Participants were chosen based on their reported frequencies of recent binge eating and drinking, with oversampling for these behaviors of interest. Participants were selected randomly from the following four constructed categories: *High Binge-High Drink (HBHD)* (binged at least three to four times per month in the past 3 months; drank alcohol on six or more occasions in the past 30 days); *High Binge-Low Drink (HBLD)* (binged at least three to four times per month in the past 3 months; drank alcohol on one or two occasions in the past 30 days); *Low Binge-High Drink (LBHD)* (no binges in the past 3 months and either a nondieter, a casual dieter, or an intense dieter; drank alcohol on six or more occasions in the past 30 days); and *Low Binge-Low Drink (LBLD)* (no binges in the past 3 months and either a nondieter, a casual dieter, or an intense dieter; drank alcohol on one or two occasions in the past 30 days).

The drinking cut-offs identified the quartile of heaviest drinkers (six or more drinking occasions in the past 30 days) and the quartile of lightest drinkers (one or two drinking occasions in the past 30 days). Since some drinking precollege is normative, we chose to exclude from our low-drinking groups participants who reported drinking no alcohol in the past 30 days. The binge eating cut-offs identified binge eaters at or near clinically significant frequency levels (at least 3–4 times/month) and nonbinge eaters. The descriptors of nondieter, casual dieter, and intense dieter come from the Dieting and Bingeing Severity Scale (DBSS; Kurth, Krahn, Nairn, & Drewnowski, 1995). These categories were used to exclude participants who used extreme weight control methods (e.g., vomiting, laxative use, "always" dieting) from the low binge eating groups. The HBHD category represented 2.5% ($n = 26$) of the pool of 1,038, the HBLD represented 1.4% ($n = 15$), the LBHD represented 11.8% ($n = 122$), and the LBLD represented 11.9% ($n = 124$). Of the 1,038 women, 72.4% ($n = 751$) did not fit into any of the four combinations of frequent and infrequent drinking and frequent and nonexistent binge eating.

A total of 45 women were recruited to participate in this study ($n = 15$ for HBHD, $n = 10$ for HBLD, $n = 10$ for LBHD, and $n = 10$ for LBLD). Since this study was part of a pilot study aimed at using IVR to more intensively gather information on binge eating and drinking behaviors, the study's sample size was determined by the number necessary to look at the relationships of these behaviors within subjects. (These results will be presented in a future report). Only participants with at least a 50% IVR response rate (i.e., at least 42 daily reports of the 84 possible) *and* completed data from the TLFB measure were included in these analyses. (Because of variability in the number of days reporting, the reports of participants with greater than 42 but less than 84 daily reports were prorated for equivalence to an 84-day period. For example, the number of binges reported was divided by the number of days reported and the result was multiplied by 84.) These restrictions reduced the number of participants for the binge eating analyses to 38 ($n = 10$ for HBHD, $n = 8$ for HBLD, $n = 10$ for LBHD, and $n = 10$ for LBLD). The participants excluded from the binge eating analyses came from the HBHD group ($n = 5$) and the HBLD group ($n = 2$). The number of participants included in the drinking analyses was also reduced to 38 ($n = 10$ for HBHD, $n =$

8 for HBLD, $n = 10$ for LBHD, and $n = 10$ for LBLD). The participants excluded from the drinking analyses came from the HBHD group ($n = 5$) and the HBLD group ($n = 2$).

Measures

Each day of the 12-week study, participants called the IVR system via a toll-free number to answer 22 to 28 questions about the previous day. The number of questions varied because, depending on the behaviors they endorsed, different follow-up questions were asked. (We created the survey so that the chains of questions participants would be gated into, depending on whether or not they endorsed a behavior, were more or less equal in length, so that if they denied engaging in a behavior, they would have only marginally fewer questions to which to respond.) Specific to this article, participants heard a female voice ask "How many times did you binge eat yesterday?" after which they would press the numeric key on the telephone keypad that corresponded to the number of binges of the previous day. Participants were also asked the questions "How many beers did you drink yesterday?", "How many standard drinks containing liquor did you drink yesterday?", and "How many glasses of wine or wine coolers did you drink yesterday?" Their responses to these questions were summed to arrive at the total number of alcoholic beverages consumed the previous day. Each question was preceded by a keyword (e.g., "binges") immediately after which responses were permitted; this allowed practiced participants to respond quickly without needing to listen to each question in its entirety. IVR technology has been used in the study of alcohol consumption among male drinkers, and has been found to provide reliable and valid self-reports (Perrine, Mundt, Searles, & Lester, 1995; Searles, Perrine, Mundt, & Helzer, 1995). To our knowledge, this is the first time this technology has been applied to self-reported binge eating or has been used by a young female adult population.

In the TLFB procedure, participants were given a calendar grid of the 12 weeks of the IVR study, which was annotated with salient, campus-wide events (e.g., home football games, holidays). The interviewer administering the measure asked them to recall as accurately as possible the number of binges and drinks consumed, day by day, over the 12-week period, using the definitions for binges and drinks of the IVR study. The TLFB technique has been used in assessing self-reported alcohol consumption, and has been found to be a reliable and valid retrospective measure for normal drinkers in college (Sobell, Sobell, Klajner, Pavan, & Basian, 1986) and in the general population (Sobell, Sobell, Leo, & Cancilla, 1988). However, TLFB methodology appears to be more precise in assessing number of drinking days than actual level of consumption (Sobell & Sobell, 1992). TLFB has not typically been applied to binge eating behavior.

Procedure

Before the study began, participants met individually with a research assistant for an hour-long session that included questionnaires on mood, personality, and expectancies, and training in how to report on their daily behaviors (e.g., binge eating and drinking) using the IVR system. Participants were given the toll-free number to call and an identification number and password (of their creation) for confidentiality. They were also given: a detailed manual describing the IVR system and the financial reinforcement schedule, a written copy of the computer-administered questions that

they would be touch-tone responding to via IVR, and a wallet-sized card with the toll-free number and the investigator's number in case of questions or problems with the system. The manual included written definitions of what constituted an eating binge and an alcoholic drink. A binge was defined, using *DSM-IV* criteria, as "an episode of eating more food than you would usually eat in a discrete period of time (e.g., 2 hours) and often accompanied by a sense of loss of control." A drink was defined as a Standard Drink: a 12-ounce can or bottle of beer, a shot of liquor, a mixed drink, a bottle of wine cooler, a 5-ounce glass of wine.

Financial incentives were provided to encourage maximum reporting. Participants received 50 cents for each daily report and a $10 bonus for each uninterrupted calendar week of reports, resulting in a maximum possible amount earned of $162. Participants were given the opportunity to get partial financial credit for days not called in by filling out a pencil and paper version of the computer-administered IVR questions ("daily worksheet") within 2 days of the missed call. Individuals who did not provide a daily report on a given day were telephoned the following day by a research assistant to ascertain any reporting problems, to encourage continued participation in the study, and to recommend the completion of a daily worksheet for the missed call.

Participants provided daily binge eating and drinking reports via IVR for 12 weeks. At the end of the 12 weeks, participants reported retrospectively on number of binges and drinks using TLFB methodology. Gathering the same information for the same 12-week time period using both IVR and TLFB permitted investigation of the agreement of the two methods and potentially the validity of TLFB for use in assessing binge eating behavior and drinking behavior.

RESULTS

IVR response rate

Study participants provided 82.4% of the total possible 3,612 data points (84 days × 43 participants) of the IVR study. (Two participants who provided less than 3 weeks of data (one gave 9 reports, the other 16), despite numerous attempts to contact them and encourage their ongoing participation, were considered nonparticipants and were dropped from the study.) This response rate is respectable albeit lower than the 93% response rate reported by Searles et al. (1995) in a study of male drinkers over the course of 112 days. Of the 43 participants, 14.0% ($n = 6$) provided data for each of the 84 days (via a combination of calls and daily worksheets), while 74.4% (n = 32) provided data for at least three fourths of the 12-week period (at least 63 daily reports). The vast majority, 90.7% ($n = 39$), provided data for at least half of the 12-week period (at least 42 daily reports); these participants' data are included in this article. We included participants with half of the daily reports because that was the lower limit required for time series analyses to be presented in a future report. In reality, most participants provided substantially more than 42 daily reports.

Of the 43 participants, 74.4% ($n = 32$) completed at least one daily worksheet during the 12-week period. Only 9.3% ($n = 4$) of the sample had worksheet data exceeding 10% of their daily reports, and 39.5% (n = 17) of the sample had worksheet data make up no more than 1% of their daily reports. Thus, it appears that the worksheet was used as an occasional, rather than a replacement, reporting means. Analyses comparing worksheet data to IVR data yielded no significant differences in terms of mean number of binges or mean number of drinks reported.

The different categories of study participants varied in their overall response rates (i.e., data points provided by a combination of IVR reports called in and daily worksheets completed). While the HBLD, LBHD, and LBLD groups all had response rates between 88 and 90%, the HBHD group had a response rate of only 70%. Furthermore, all of the study participants who supplied less than 50% IVR data points (i.e., called in or completed daily worksheets for fewer than 42 days) came from the HBHD group.

There were also differential response rates across days of the week. While the response rate from Sunday through Thursday averaged from 82 to 88%, the response rate for the weekend days of Friday and Saturday averaged 77% and 79%, respectively.

Comparisons were made between study participants who provided sufficient IVR data ($n = 39$) and those who did not ($n = 6$) (i.e., who provided data for fewer than 42 out of 84 days) on the eating- and drinking-related screening (baseline) data. The "completers" and the "drop-outs" differed on items from both the eating and drinking domains. Drop-outs were more likely to have reported engaging in any of several purging behaviors over the past 3 months. In particular, compared to completers, drop-outs were 6.4 times, $t(42) = 7.39$, $p < .01$, more likely to have vomited, 7.2 times, $t(42) = 4.38$, $p < .01$, more likely to have used laxatives, 7.5 times, $t(43) = 10.61$, $p < .01$, more likely to have fasted, and 2.8 times, $t(44) = 2.14$, $p < .05$, more likely to have endorsed often or always dieting. In terms of eating and body attitudes, drop-outs were 3.3 times, $t(44) = 3.50$, $p < .01$, more likely than completers to report rarely or never feeling satisfied with their body shape, and 6.0 times, $t(44) = 5.48$, $p < .01$, more likely to often or always feel out of control when eating. Odds ratios could not be created for the likelihood of endorsing having "ever binged" because all of the six drop-outs endorsed this behavior (compared to 20 of the 39 completers). The groups differed on drinking items as well. Compared to completers, drop-outs were 6.9 times, $t(43) = 11.47$, $p < .01$, more likely to have endorsed the highest categorical frequency of lifetime drinking (40 or more occasions), and 3.3 times, $t(44) = 3.50$, $p < .01$, more likely to have drunk on more than two occasions per week for the past month. Drop-outs were also 6.7 times, $t(44) = 9.43$, $p < .01$, more likely to have reported drinking five drinks in a row on at least three occasions in the past 2 weeks. Odds ratios could not be created for the likelihood of drinking to "feel drunk" because all of the six drop-outs endorsed drinking on most or nearly all occasions to feel drunk (compared to 15 of the 39 completers).

Binge eating

Figure 1 shows the distribution of the average binge eating frequency across the two methods (IVR and TLFB) for the same reporting period of 12 weeks (our first objective). Inspection of the data in this figure reveals that while 44.7% ($n = 17$) of the participants reported no binge eating using the IVR method, 60.5% ($n = 23$) reported no binge eating using the TLFB method. At the same time, 15.8% ($n = 6$) of the participants reported binge eating on average twice per week using the IVR method, while 7.9% ($n = 3$) reported binge eating with that frequency using the TLFB method. TLFB reports of binge eating were 1.9 times, $t(37) = 2.80$, $p < .01$, more likely than IVR reports to endorse no binge eating, and IVR reports were 2.2 times, $t(37) = 1.68$, $p = 0.1$, more likely than TLFB reports to endorse binge eating frequencies of twice or more weekly. The difference in the reports of twice-weekly binge eating is especially important for researchers and clinicians interested in identifying people meeting criteria for a diagnosis of bulimia nervosa or binge eating disorder.

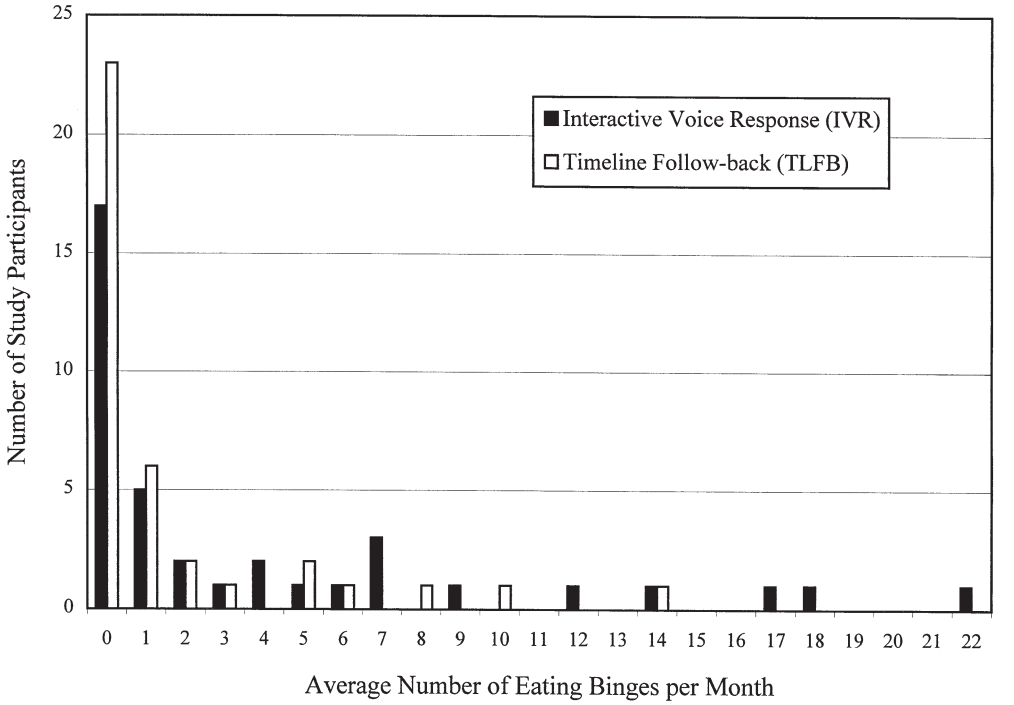

Fig. 1. Distribution of average number of eating binges per month across two methods (IVR, TLFB) for the same reporting period of 12 weeks.

When limiting our inspection of the level of agreement of binge eating reports to only the 22 participants who reported binge eating via at least one method, we found discrepancies. Seven of the 22 participants (31.8%) who indicated some binge eating reported no binge eating with the TLFB but some binge eating (i.e., at least once a month) with the IVR. Furthermore, 13 of the 22 binge eating participants (59.1%) had TLFB reports that were less than 50% of their IVR reports of binge eating.

Our second objective was to calculate indices of sensitivity and specificity for the TLFB methodology in terms of its identification of individuals who binge eat two or more times per week, which *DSM-IV* identifies as an important diagnostic cut-off for bulimia nervosa and binge eating disorder. The sensitivity of a test is the test's potential for making a correct diagnosis in confirmed positive cases. That is, it is an index of the test's ability to find an effect when it is there. For example, a method has good sensitivity if it identifies as twice-weekly binge eaters, individuals who really do binge eat two or more times per week according to the daily IVR reports, the gold standard for this study. The specificity of a test is the test's potential for making correct diagnoses in confirmed negative cases. That is, it is an index of the test's ability to not find an effect if it is not there. For example, a method has good specificity if it correctly identifies individuals who binge eat less than twice a week. The TLFB had high specificity (1.0) but poor sensitivity (0.33) for identifying binge-eating rates of two or more times per week. That is, TLFB accurately identified non- or low-frequency binge eaters, but failed to identify a substantial portion of twice-weekly binge eaters as such.

Drinking

Figure 2 shows the distribution of the average number of drinks per month across the two methods (IVR and TLFB) for the same reporting period of 12 weeks (our first objective).

In terms of the criterion of seven or more drinks per week, both IVR and TLFB identified 14 participants meeting the criterion, 11 of whom were identified by both methodologies. In terms of the criterion of five or more drinks in a row at least three times in a 2-week period on average, IVR identified seven participants meeting the criterion, while TLFB identified eight; the methodologies identified six participants in common. Thus, we found convergence between IVR and TLFB for drinking reports; the methods generally identified the same participants.

In terms of our second objective of indices of sensitivity and specificity for problem drinking, the TLFB appeared to have good sensitivity (0.79) and good specificity (0.88) for identifying participants drinking seven or more drinks per week, as well as good sensitivity (0.86) and good specificity (0.94) for identifying participants drinking five or more drinks in a row at least three times in a 2-week period. That is, the TLFB accurately identified both problem drinkers and non-problem drinkers, whether the drinking measure was number of drinks or number of binge drinking occasions.

DISCUSSION

The main finding of this study is that while TLFB appeared to fairly accurately measure drinking behavior, this same methodology did a poor job of representing binge eating. In assessing binge eating, TLFB data were discrepant from actual binge eating behavior as recorded by daily IVR. In assessing drinking, TLFB data mapped on well to actual drinking behavior as recorded by daily IVR.

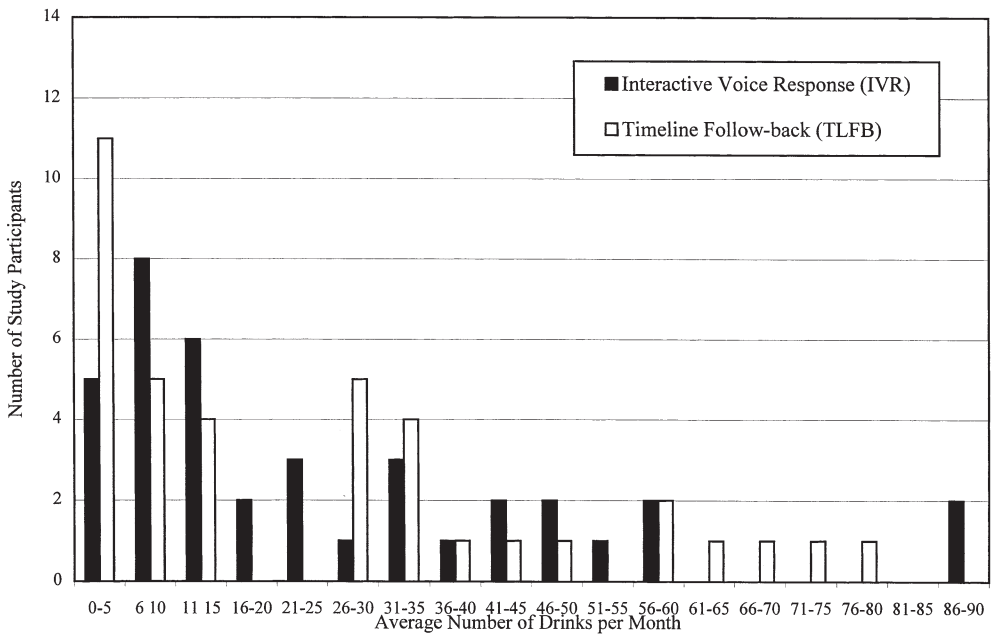

Fig. 2. Distribution of average number of drinks per month across two methods (IVR, TLFB) for the same reporting period of 12 weeks.

In going beyond how well the methods chart the behaviors, to how well they identify behavior of potential clinical concern, a similar pattern emerges. In identifying participants with binge eating frequencies of twice or more weekly, TLFB identified only a small subset of these frequent binge eaters, providing a conservative estimate that misses true cases. In identifying participants with problem drinking (defined by weekly average or frequency of binge drinking), there was considerable overlap between the TLFB reports and daily IVR reports, such that these methods tended to select the same problem drinkers.

Indices of sensitivity and specificity provide further evidence for the differential validity of TLFB methodology across target behaviors. In measuring binge eating behavior of 2 or more times per week, the TLFB was specific but not sensitive in its identification of this behavior. In measuring problem drinking behavior, defined as either seven or more drinks per week or five or more drinks in a row at least three times in a 2-week period, TLFB was both sensitive and specific. Again, TLFB appears to accurately identify problem drinking, but not the diagnostic criterion of twice-weekly binge eating.

Why is it that TLFB methodology is adequate for drinking assessment but woefully inadequate for binge eating assessment? We speculate that, since the TLFB is administered in the view of an investigator, the answer lies in the shame attached to the behavior. In this population (female freshmen attending a large public university) there may be little reason for reserve or incomplete disclosure when it comes to reporting drinking behavior. If drinking, and even binge drinking, on the college campus is commonplace, "admitting" to drinking eight drinks in a night may not be shameful, so there would be no impetus to underreport. Thombs, Wolcott, and Farkash (1997) found that, in a college population, many young people perceive heavy drinking as normative and that perceptions of close friends' drinking intensity is related to consumption levels. Binge eating, however, does not have the same developmentally normative status as drinking. Hence, women may be more likely to underreport binge eating behavior than drinking behavior when in the presence of an investigator, as is the case using TLFB methodology, because of the ignominy attached to binge eating.

This study provides some support for the viability of the IVR technology for the assessment of binge eating and drinking behaviors in a female college population. While the response rate attained was respectable (82.4%), it was not as high as that attained by Searles et al. (1995) (93%) among their group of adult men. There are several possible explanations for this difference in responding. It is possible that college students, especially those in their first year of adjustment to the new environment, present unique challenges to assuring daily compliance. For example, the underreporting on the weekend seen in this study may be unique to the college population, given the dramatic increase in drinking opportunities during the weekend compared to the weekdays. Also, our financial incentive may not have been as motivating to a group of college students, perhaps financially supported by their parents, as to Searles' group of independently living men. Another way to explain this study's lower response rate is to look at the differential response rates across the groups selected on the basis of binge eating and drinking frequencies. It appears that the overall response rate in this study was adversely affected by the lower response rate (70%) of the participants selected for having both frequent binge eating *and* frequent drinking (HBHD). The lower response rate of the HBHD group was largely due to their having the highest drop-out rate. In comparison, participants reporting frequent binge eating *or* frequent drinking *or* neither at baseline (HBLD, LBHD, LBLD) had approximately a 90% re-

sponse rate. Future studies applying IVR technology to the study of binge eating and drinking behaviors in a female college sample should be aware of the added difficulties of getting daily reports from college students. Oversampling groups with multiple disinhibited behaviors (e.g., binge eating *and* drinking) should also be considered since members of this group are most likely to drop-out, yielding a lower response rate for the group.

Several novel advances come with using IVR technology to provide a more intensive and on-line study of behaviors of interest. First, recall time is reduced to 24 hours, thus avoiding the pitfalls of forgetting that can occur when participants are asked to recall something that happened days, weeks, or even months ago. This feature of IVR, along with the anonymity of the methodology, bolsters this study's claim that data provided using IVR are most representative of actual behavior. Second, participants' touch-tone responses are directly entered into a database, thus minimizing transcription errors and financial cost. Third, the flexibility of IVR technology (i.e., being able to call the toll-free number regardless of a participant's geographic location in the United States) permits participants to maintain consistent reporting while travelling. Fourth, the minimum amount of time required of the participants permits the incorporation of the reporting into participants' normal daily routines with little effort. Participants spent on average 1.87 minutes per day ($SD = 0.21$ minutes) on touch-tone reporting, with a range of 1.59 minutes to 3.00 minutes. Their average times decreased along the course of the study period as they became familiar with the questions asked and the mechanics of touch-tone responding. Although even simpler and less expensive methods are certainly available, IVR technology is important for reasons beyond accuracy. For example, by using IVR technology to ask a set of questions on domains hypothesized to be related (e.g., questions on mood, stressors, binge eating, and drinking), researchers can obtain the details necessary to understand the dynamics of different behaviors, feelings, and thoughts at a microlevel.

Although this study has many strengths, there are also limitations to consider. For example, the very process of daily reporting may have implications for ease of recall in retrospective reports. It is possible that the self-monitoring required in the daily report protocol may have increased participants' awareness of their behaviors and may have inflated the similarity between retrospective reports and actual behavior. However, the finding that the TLFB and IVR data did not agree well for binge eating provides confidence that participation in the IVR portion of the study did not automatically influence TLFB response toward higher convergence. Future research should consider the effects that daily reporting (i.e., use of IVR) may have on retrospective recall. Another limitation is that the "drop-outs" in this study had more severe indices of experience with binge eating, purging behavior, eating attitudes, and drinking behavior at baseline than the "completers." Thus, the individuals most frequently endorsing these problem behaviors did not provide data for analysis. Attrition of participants with problem behaviors is a common problem in any longitudinal study, and unfortunately, using IVR technology does not appear to provide immunity. It may be that individuals with multiple or more severe problem behaviors find it difficult to delay gratification, and thus are not effectively maintained in a study with a financial incentive system that delays gratification. In that case, one possible solution would be to set up the financial reward system so as to provide more immediate gratification for study participation (e.g., payment daily or weekly instead of at the end of the study). Nonetheless, the majority of participants with frequent binge eating and/or frequent drinking at baseline completed the study. Finally, cases of not reporting in this study may have

been due to the very behaviors of interest: drinking and binge eating. Although it is possible that some of the missing days were due to relevant behavior (e.g., a drinking binge), participants did call in to report these behaviors.

In conclusion, this study found that TLFB data well-represented actual drinking behavior, as represented in the IVR data, but underrepresented actual binge-eating behavior. Alcohol researchers, comfortable with and confident in TLFB as an assessment tool for drinking behavior, may be tempted to use TLFB methodology to also capture binge eating behavior, if they are interested in issues of comorbidity and multiple addictions. Our findings suggest that there are significant problems in applying this methodology, successful in the study of drinking behavior, to the study of binge eating: TLFB data did not accurately represent binge eating behavior as captured by the daily anonymous IVR reporting. This study also made use of a new methodological assessment tool, IVR technology, and demonstrated that it is a viable research methodology in the study of binge eating and drinking among a young female adult population, although the uniqueness of the college context and the presence of multiple disinhibited behaviors must be taken into account for maximum success.

REFERENCES

Adinoff, B. (1998). *Alcohol*. Lecture presented at the American Academy of Addiction Psychiatry Review Course, Kansas City, MO.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Johnston, L. D., O'Malley, P. M., & Bachman, J. G. (1994). *National survey results on drug use from the Monitoring the Future Study, 1975–1993: Vol. 1. Secondary school students*. Rockville, MD: U.S. Department of Health and Human Services.

Kurth, C. L. (1994). *The relationship between dieting severity and alcohol use in young women*. Unpublished doctoral dissertation, University of Minnesota.

Kurth, C. L., Krahn, D. D., Nairn, K., & Drewnowski, A. (1995). The severity of dieting and bingeing behaviors in college women: Interview validation of survey data. *Journal of Psychiatric Research*, **29**, 211–225.

Perrine, M. W., Mundt, J. C., Searles, J. S., & Lester, L. S. (1995). Validation of daily self-reported alcohol consumption using interactive voice response (IVR) technology. *Journal of Studies of Alcohol*, **56**, 487–490.

Searles, J. S., Perrine, M. W., Mundt, J. C., & Helzer, J. E. (1995). Self-report of drinking using touch-tone telephone: Extending the limits of reliable daily contact. *Journal of Studies of Alcohol*, **56**, 375–382.

Sobell, L. C., & Sobell, M. B. (1992). Timeline follow-back: A technique for assessing self-report alcohol consumption. In R. Z. Litten & J. P. Allen (Eds.), *Measuring alcohol consumption: Psychosocial and biochemical methods* (pp. 41–72). Totowa, NJ: Humana Press.

Sobell, M. B., Sobell, L. C., Klajner, F., Pavan, D., & Basian, E. (1986). The reliability of a timeline method for assessing normal drinker college students' recent drinking history: Utility for alcohol research. *Addictive Behaviors*, **11**, 149–161.

Sobell, L. C., Sobell, M. B., Leo, G. I., & Cancilla, A. (1988). Reliability of a timeline method: Assessing normal drinkers' reports of recent drinking and a comparative evaluation across several populations. *British Journal of Addiction*, **83**, 393–402.

Thombs, D. L., Wolcott, B. J., & Farkash, L. G. (1997). Social context, perceived norms and drinking behavior in young people. *Journal of Substance Abuse*, **9**, 257–267.