Alcohol Consumption Measures

Linda C. Sobell, Ph.D., ABPP, and Mark B. Sobell, Ph.D., ABPP

Nova Southeastern University, Ft. Lauderdale, FL

Contributions from markedly different kinds of studies—biogenetic, epidemiologic, longitudinal, population surveys, clinical analog, and treatment outcome—have advanced our understanding of alcohol use and abuse. Although different studies examine issues from different perspectives, they have one thing in common—the assessment of alcohol consumption. Alcohol consumption, however, is a complex behavior that can change considerably over time.

Twenty-five years ago very few drinking measures existed. Today the situation has changed dramatically (Alanko 1984; Room 1990; L.C. Sobell and Sobell 1992, 1995). Multiple measures are now available. Thus, the issue now is how to select the best measure for a given purpose, as each measure has advantages and limitations. This chapter, like most in this *Guide*, was first published in 1995 (L.C. Sobell and Sobell 1995). This update reviews the literature on drinking measures published through mid-2001, presents new measures that met the inclusion criteria for this volume, and provides recommendations about what drinking measures to use and for what purpose.

When selecting a drinking measure, a decision must be made about the type of information needed (e.g., level of precision, timeframe, amount of time that can be devoted to data collection). Ultimately, the utility of a drinking measure for research and/or clinical purposes will rest on its intended use. Therefore, the following questions need to be answered when selecting a drinking measure:

- How is the information to be used?
- Over what time interval should data be collected?
- How long will it take to collect the data?
- What type of drinking information (e.g., precision) is needed?
- What level of training or expertise is needed to administer the instrument?
- Is the measure psychometrically reliable and valid?

Another critical but often overlooked issue is the interviewer's role. The ease with which respondents complete drinking measures depends partly on the interviewer's attitude. The interviewer's familiarity with the method and with techniques to elicit recall will not only facilitate completion of the measures but will also ensure more accurate data collection.

SELF-REPORT ISSUES

Because the assessment and evaluation of drinking is largely dependent on self-reports, validity and reliability are important issues. The primary issue is whether such reports are accurate. Several reviews of the validity and reliability of self-reports of drinking have been published, so only selected issues will be addressed in this chapter, and then only briefly. Interested readers should consult the reviews noted in this chapter for in-depth discussions.

Clinic Populations

Most information from alcohol abusers in research and clinical settings, whether for diagnostic, assessment, treatment, or outcome purposes, comes from clients (Del Boca and Noll 2000). Consequently, the alcohol field is greatly dependent on self-reports. Several comprehensive reviews of the validity and reliability of alcohol abusers' self-reports have concluded that self-reports are generally accurate and can be used with confidence if the data are gathered under specific conditions (Babor et al. 1990; Maisto et al. 1990; L.C. Sobell and Sobell 1990; Brown et al. 1992; Babor et al. 2000). Factors shown to enhance accurate self-reporting include when people are (a) alcohol free when interviewed; (b) given written assurances of confidentiality; (c) interviewed in a setting that encourages honest reporting (e.g., clinical or research versus probation office); (d) asked clearly worded objective questions (e.g., "How many times have you been arrested for drunk driving?") versus subjective questions (e.g., "Did you get drunk last night?"); and (e) provided memory aids (e.g., calendar for aiding recall of drinking).

With one or two exceptions, these reviews have shown that alcohol abusers usually describe themselves more negatively (i.e., more heavy drinking and related consequences) than does data from other sources (e.g., reports from collaterals or liver function tests). There is one condition, however, when alcohol abusers' self-reports tend to underestimate consumption—when they are interviewed with any alcohol in their system (L.C. Sobell and Sobell 1990; L.C. Sobell et al. 1994). Interestingly, alcohol abusers also report that their self-reports would be most accurate when they are alcohol free, and that their self-reports would likely be increasingly inaccurate as a function of the amount of alcohol they had consumed (L.C. Sobell et al. 1992). One way to ensure that people are alcohol free when interviewed is to use a breath tester to assess alcohol use before the interview (L.C. Sobell et al. 1994); several inexpensive portable breath testers are available. It should be noted that therapists' judgments about clients' level of drinking are frequently inaccurate (M.B. Sobell et al. 1979), probably because of the phenomenon of tolerance.

A sizable body of literature clearly demonstrates that as a group alcohol abusers' self-reports of their drinking and related consequences can be used with confidence (Schwarz 1999; Babor et al. 2000; Del Boca and Noll 2000). While some small proportion of alcohol abusers' self-reports in each study will be inaccurate, currently, with a few exceptions, it is difficult to identify individuals who give inaccurate self-reports (reviewed in Toneatto et al. 1992). Two conditions, however, that are predictive of less consistent self-reports are (a) alcohol abusers who report a long drinking history (i.e., years problem drinking) (Toneatto et al. 1992; Drake et al. 1995; Babor 1996) and (b) questions that require a subjective judgment (i.e., difficult to define or ambiguous) (see Toneatto et al. 1992).

Survey Studies

Reports of drinking in population surveys have shown bias in terms of aggregate consumption. When projected to the total population, for example, this bias only accounts for a portion of total beverage sales (reviewed in Midanik 1982; Poikolainen and Kärkkäinen 1985). Several explanations have been offered regarding why alcohol consumption is underreported in general population surveys (Midanik 1982; Alanko 1984; Lemmens and Knibbe 1993; Göransson and Hanson 1994):

- Heavy drinkers have a high nonparticipation rate in surveys.
- Forgetting increases with increasing consumption.
- The study method may be prone to bias. For example, quantity-frequency (QF) measures (estimates of average quantity and frequency; see the "Review of

Drinking Measures" section of this chapter) result in greater underestimates than daily diaries.

- Questionnaire construction may affect responses (e.g., questionnaires with more questions about atypical drinking result in reports of greater consumption).
- Timeframe may affect response (e.g., seasonal variation affects estimates).

Several studies show that with minimal sampling problems and heavy drinking factored into aggregate consumption, the variability between reports of drinking and alcoholic beverage sales figures can be substantially reduced (Midanik 1982). A report describing two Swedish alcohol surveys sheds some light on discrepancies between reports gathered using different methods (Kuhlhorn and Leifman 1993). Both surveys were conducted by respected research groups and used large numbers of respondents. The two surveys yielded very large differences in their retail sales coverage rates (i.e., registered alcoholic beverages sales), namely, 75 percent and 28 percent. In the survey with a high coverage rate, respondents' daily drinking patterns were able to reflect heavy drinking on weekends by dividing a "normal week's" drinking into four periods (Monday–Thursday, Friday, Saturday, and Sunday). Because the survey with a low coverage rate used a QF measure, a "normal week's" drinking could not be similarly derived. A test of internal validity of the survey with the higher coverage confirmed that the increased coverage was due to the refined nature of the questions.

Other surveys using heavy or atypical drinking questions have reported similar increases in estimates of alcohol consumption. Polich and his colleagues (Polich and Orvis 1979; Armor and Polich 1982) used an adjusted QF method that asked for typical and atypical drinking and found that, by adding questions about heavy drinking days, there was a 43 percent increase in daily per capita consumption. In a study by Göransson and Hanson (1994), while 15.1 percent of consumers increased their reported drinking using an adjusted QF measure, the overall change in weekly per capita consumption was modest.

Some survey studies have used a recent drinking occasions measure (also called the Finnish period estimate method) or a situation-specific measure (Mäkelä 1971; Hilton 1986; Midanik 1994; Single and Wortley 1994; Wyllie et al. 1994). Such measures ask respondents to report their alcohol use over a time interval involving a number of drinking occasions or specific drinking situations. For each measure, several variations are possible, and, as one might expect, studies using different variants have resulted in different amounts of alcohol reported consumed.

Self-Report Summary

The literature suggests that although the accuracy of an individual's report may be difficult to determine, from a group perspective self-reports of alcohol use from clinical and nonclinical samples are accurate when people are interviewed under the conditions discussed earlier. In addition, it appears that questions about heavy or atypical drinking must be included to accurately capture a person's total alcohol consumption.

REVIEW OF DRINKING MEASURES

Although a number of drinking measures have been developed and reported in the literature, only five satisfied the criteria for inclusion in this *Guide*. Tables 1A and 1B provide descriptive and administrative information for these five measures; see the fact sheets in the appendix for more detail. Table 2 lists how each of these five measures has been psychometrically evaluated. Four of the measures assess drinking only; Form 90 also assesses domains other than alcohol use.

All five measures have been used with adults and adolescents. Most have been used with clinical

TABLE 1A.—Drinking measures: Descriptive information

Measure	Purpose	Clinical utility	Drinking variables generated	Assessment timeframe	Target population	Groups used with
TLFB	Assessment of daily drinking; several dimensions of drinking can be separated and examined	Individual picture of main features of past drinking; advice and feedback during treatment; monitoring progress	Daily drinking into user-defined categories, variability, pattern, level of drinking, time to first relapse	Retrospective recall of 30–360 days before interview	Adults and adolescents	Alcohol abusers and normal drinkers; males and females; college students
Form 90	Assessment of daily drinking using a calendar and weekly grid	Individual picture of main features of drinking in the 90 days before last drink	Same as for TLFB except uses a 90-day interval before last drink	Retrospective recall of 90 days before last drink	Adults and adolescents	Alcohol abusers; males and females
DSML	Daily recall of drinking	Advice and feedback during treatment; monitoring progress	Same as for TLFB	Recall of daily drinking	Adults and adolescents	Alcohol abusers and normal drinkers; males and females; college students
LDH	Chronological recall of drinking patterns from adolescence to adulthood	Information about lifetime drinking patterns	QF average and maximum of drinking phases	Retrospective lifetime assessment of drinking	Adults and adolescents	Alcohol abusers and normal drinkers
QF measures[1]	Assessment of drinking	Rapid information about number of days drinking and overall consumption	QF, QF volume, volume variability	Retrospective recall of typical month or last 30 days	Adults and adolescents	Alcohol abusers and normal drinkers; college students

Note: The measures are listed in the same order in which they are discussed in the text; see the text for the full names of the measures.
[1] Individual QF measures are summarized in table 3.

TABLE 1B.—Drinking measures: Administrative information

Measure	Languages other than English	Administration options	Time to administer (min.)	Scoring time (min.)	Training needed?	Computerized version?[1]	Fee for use?
TLFB	Belgian Dutch, Belgian French, French, German, Japanese, Polish, Spanish, Swedish	P&P, interview, computer	10–15 for 90 days, 30 for 360 days	10 for P&P, 5 for computer	Yes	Yes	No for P&P, yes for computerized version
Form 90	None	Interview	40–60 (assessment version)	20	Yes	No	No
DSML	Japanese, Polish, Spanish, Swedish	P&P	NA	NA	No	No	No
LDH	None	Interview	20–30	5–10	Yes	No	No
QF measures[2]	Spanish	P&P	4–60	5	No	No	No

Note: The measures are listed in the same order in which they are discussed in the text; see the text for the full names of the measures.

NA = not applicable; P&P = pencil and paper.

[1] Computer version of the measure, not computerized scoring.

[2] Individual QF measures are summarized in table 3.

TABLE 2.—Availability of psychometric data on drinking measures

Measure	Reliability		Validity		
	Stability	Internal consistency	Content	Criterion	Construct
TLFB	•		•	•	•
Form 90	•			•	•
DSML	NA		•		•
LDH	•		•		•
QF measures	•		•	•	•

Note: The measures are listed in the same order in which they are discussed in the text; see the text for the full names of the measures. NA = not applicable.

and normal drinker populations and evaluated with males and females. The five drinking measures can be classified into one of two general recall methods: (a) *Quantity-Frequency:* retrospective estimates of average daily consumption and the average frequency with which consumption occurs; and (b) *Daily Drinking:* retrospective estimates of drinking that occur on each day in the interval.

Four of the five measures collect retrospective data (i.e., information about alcohol use after it occurs). The one concurrent measure, Drinking Self-Monitoring Log (DSML), asks people to record their drinking at about the same time as it occurs. The assessment timeframe over which the measures obtain data range from daily recall, to retrospective recall of drinking in the past year, to lifetime drinking. Not all of the measures inquire about a specific interval; some ask about a "typical" period. Only one of the drinking measures is available in a computerized format. With respect to administration time, the measures vary from about 5 minutes for a brief QF measure, to 30 minutes for a 12-month Timeline interview, to 40–60 minutes for Form 90. Time to score the measures is relatively short (i.e., 5–20 minutes). Some training is required for administration of all of the measures. All pencil-and-paper versions of the measures are available for use without charge.

The summaries presented below will help readers select a measure best suited for their purpose (see the fact sheets in the appendix to this *Guide* for more detail). Selecting a drinking measure requires consideration of several factors: population, time available for the assessment, how the information will be used, timeframe of reports, and the types of information needed. While day-by-day precision cannot be assumed or necessarily expected with any measure, some measures will provide a more complete picture of a person's drinking than others will.

Alcohol Timeline Followback

The Alcohol Timeline Followback (TLFB), a daily drinking estimation method, provides a detailed picture of a person's drinking over a designated time period. The TLFB method was originally developed as a research tool for use with alcohol abusers, but it has since been adapted for use in clinical settings and has been extended to measure drug and cigarette use (L.C. Sobell et al. 1994; L.C. Sobell and Sobell 1995, 2000). The TLFB has been extensively evaluated with a wide range of clinical and nonclinical populations (L.C. Sobell and Sobell 1992, 1995, 2000) and was chosen by the American Psychiatric Association

as having met criteria for inclusion in their *Handbook of Psychiatric Measures* (American Psychiatric Association 2000).

The TLFB is a calendar-based form in which people provide retrospective estimates of their daily drinking, including abstinent days over a specified period of time ranging up to 12 months prior to the interview. Memory aids are used to enhance recall. The amount of time needed to administer the TLFB varies as a function of the assessment interval (e.g., 90 days = 10–15 minutes; 12 months = 30 minutes).

The TLFB can generate a number of variables that provide more precise and varied information about a person's drinking than is produced by QF methods. The TLFB can generate variables to portray pattern, variability, and level of drinking. Administration of the TLFB is flexible: It can be self-administered or administered in person by trained interviewers, and it is available in pencil-and-paper and computerized formats (L.C. Sobell and Sobell 1996*a*). It has been translated into French, German, Japanese, Polish, Spanish, and Swedish. The TLFB can collect drinking data for different purposes (i.e., assessment, followup, and collateral followup) and by multiple methods (i.e., in person or by phone, mail, or computer). Finally, the TLFB has very good psychometric characteristics with a variety of drinker groups.

Daily drinking recall methods and retrospective daily diaries that are like the TLFB method have been used in other studies with similar results (Redman et al. 1987; Werch 1989; Flegal 1990; Webb et al. 1990; O'Hare 1991; O'Hare et al. 1991; Webb et al. 1991; Lemmens et al. 1992). The TLFB was adapted for use in Project MATCH (Miller and Del Boca 1994; Tonigan et al. 1997), a multisite matching trial of the National Institute on Alcohol Abuse and Alcoholism (NIAAA). This adaptation, called Form 90, assesses drinking as well as other domains and is discussed later in this chapter.

Alcohol Timeline Followback (TLFB)

RECOMMENDED USE: To evaluate specific changes in drinking. Use when relatively precise estimates (versus QF methods) of drinking are necessary, especially when a complete picture of the distribution of drinking days (i.e., high- and low-risk days) is needed.

ADVANTAGES: This is the measure of choice when drinking is variable (e.g., problem or binge drinkers), or when relatively precise estimates of drinking are needed (e.g., frequency of drinking at specific levels). The pattern, variability, and level of drinking can be profiled using variables such as percentage of days drinking at different levels or the pattern of weekend/weekday drinking.

A discussion of the results of the TLFB with the client can be used to point out triggers to use, high-risk situations, and relapse periods. Repeated administrations of the TLFB from assessment, over the course of treatment, and throughout followup will produce a continuous profile of changes in drinking. The profile can assist clients in seeing where progress has been made and where problems still exist. A video is available to train interviewers in how to use this method (L.C. Sobell and Sobell 1996*b*).

The TLFB can be used in treatment as an advice-feedback tool. For example, using the information provided by a client on the TLFB, a personalized feedback summary that includes group norm comparisons of the person's drinking in the past year as well as health risk indicators and the cost of drinking can be prepared. Such feedback can be used to enhance a client's motivation and increase commitment to change (L.C. Sobell et al. 1996; Treatment Improvement Protocol Series 35 Consensus Panel 1999).

LIMITATIONS: If time is at a premium or less precise information about drinking is needed (e.g., some survey studies), the TLFB would be too

demanding. In addition, in some situations (e.g., mailed-out questionnaires) the TLFB may not be justified because it increases the burden on respondents, which may in turn result in increased attrition rates (Cunningham et al. 1999; L.C. Sobell et al. in press). In such cases, a QF measure can increase the percentage of clients for followup, albeit with less specific drinking data (L.C. Sobell et al. in press).

Form 90

Form 90 can generate baseline and followup information. Besides collecting daily drinking information for 90 days prior to the last drink, Form 90 also collects data on other aspects of clients' functioning (e.g., use of drugs; experience with medical and psychological treatments; lifestyle activities such as work, school involvement, and religious participation). Form 90, which was developed for Project MATCH (1993),

> combined two previously published methods for assessing alcohol consumption. A calendar base is used to ensure a continuous record for each day in the assessment period, in the manner of the Timeline Followback (TLFB) method ([L.C.] Sobell and Sobell 1992). Because drinking patterns often manifest consistency from week to week or from episode to episode, a grid averaging method (Miller and Marlatt, 1984) was incorporated to capture efficiently such consistent patterns when they occur, inserting them into appropriate sections of the calendar (Tonigan et al. 1997, p. 358).

Form 90 has been shown to have convergent validity with QF and grid measures (Grant et al. 1995) as well as satisfactory reliability "when interviewers have received careful training and supervision in its use" (Tonigan et al. 1997, p. 358). Form 90 can be used to collect drinking data

for various applications (i.e., intake; telephone followup; collateral intake and followup).

Form 90

RECOMMENDED USE: To evaluate specific changes in drinking before and after treatment for 90 days before the date of the last drink. Use when relatively precise estimates of drinking are needed.

ADVANTAGES: When drinking is variable (e.g., problem or binge drinkers) or when relatively precise estimates of drinking are needed (e.g., frequency of drinking at specific levels). The pattern, variability, and level of drinking can be profiled using variables such as percentage of days drinking at different levels or the pattern of weekend/weekday drinking. Assessment data from Form 90 can be used in treatment as an advice-feedback tool to enhance a client's motivation to change.

LIMITATIONS: If time is at a premium or less precise information about drinking is needed (e.g., survey studies or physicians' offices), Form 90 would be too demanding because it takes 40–60 minutes to obtain 90 days of drinking and related information. Although Form 90 can collect sequential 90-day chunks of drinking data, its psychometric evaluation has been limited to the 90 days before the date of the last drink. Thus, if more than 90 days are needed (e.g., comparable 1-year pretreatment and 1-year posttreatment data), then the TLFB method should be used because it has good psychometric characteristics for daily drinking data up to 360 days from the date of the interview. In addition, Form 90 cannot be used in some situations (e.g., mailed-out questionnaires, surveys, and self-help interventions) because the authors feel it requires trained interviewers.

Drinking Self-Monitoring Log

Self-monitoring of drinking involves recording consumption on a daily or a drink-by-drink basis.

In contrast to other measures in this chapter, which are retrospective, self-monitoring is intended to concurrently record different aspects of alcohol use (e.g., amount, frequency, mood, urges) when it occurs. Self-monitoring has been widely used for assessment and treatment monitoring of different behaviors (Korotitsch and Nelson-Gray 1999). With respect to alcohol use, several logs and diaries have been used over the years (Vuchinich et al. 1988; L.C. Sobell et al. 1994). Because drinking is recorded either when it occurs or shortly thereafter, this method is subject to fewer memory problems than retrospective measures (Samo et al. 1989; M.B. Sobell et al. 1989; Lemmens et al. 1992). That is, slightly higher frequency of drinking is reported by DSML than by retrospective methods, although reports of amount consumed per drinking day are not affected by method type. One limitation, however, is that not all individuals comply with self-monitoring instructions (Sanchez-Craig and Annis 1982).

An important issue with assessing drinking concurrently is that while accuracy might improve, recording one's drinking may be reactive (i.e., the method of recording may impact drinking by reducing it) and could seriously confound research designs. Not only is the evidence for the reactivity of self-monitoring weak, but few studies have used clinical populations (Nelson and Hayes 1981; Korotitsch and Nelson-Gray 1999). In two clinical trials where self-monitoring was used as a control/waiting condition, significant reductions in drinking were observed (Harris and Miller 1990; Kavanagh et al. 1999). It should be noted, however, that for clinical purposes, reactivity may be desirable (e.g., feedback is intended to encourage clients to reduce their drinking).

Drinking Self-Monitoring Log (DSML)

RECOMMENDED USE: When slightly more accurate information about the frequency of drinking is necessary or desired, and for obtaining reports of daily drinking reports during treatment.

ADVANTAGES: Self-monitoring provides feedback about treatment progress and can be used to identify situations that pose a high risk of relapse (e.g., monitoring urges); it also gives clients an opportunity to discuss their drinking during treatment. When used during treatment in conjunction with a retrospective daily recall method, self-monitoring provides a continuous record of daily drinking from pretreatment throughout treatment. Discussion of self-monitoring during treatment gives clients advice and feedback about changes in their drinking and related behaviors (e.g., urges, avoidance of high-risk situations) and allows them to evaluate their progress toward their goals. Such advice can enhance or strengthen motivation for change. For example, if positive changes have occurred, discussion of these changes can be used to increase a client's self-efficacy (e.g., "That is a big change from when you entered treatment. How were you able to not drink this past week?").

LIMITATIONS: Because self-monitoring cannot provide retrospective drinking data, it can only be used for pretreatment assessment if a baseline monitoring period precedes treatment. In addition, some individuals will not comply with instructions to self-monitor their drinking.

Lifetime Drinking Measures

Measures of lifetime drinking structurally parallel QF methods because they ask about average quantities and average frequencies of drinking, but over an entire drinking career or very long time periods (L.C. Sobell et al. 1993). Three different lifetime drinking measures have been developed. The first and most widely used, the Lifetime Drinking History (LDH) (Skinner and Sheu 1982), is a structured QF measure that captures distinct phases and changes in a person's lifetime

drinking patterns by asking about the typical and maximum quantity consumed per occasion as well as the frequency of drinking for the typical and maximum amount. Because the LDH allows respondents to report their own temporal phase changes, it uses a floating time interval to collect data. The goal is to obtain information about people's alcohol use over their drinking career, specifically capturing major changes in drinking patterns. To better capture frequent heavy drinking patterns, a maximum frequency category was added to the original LDH (L.C. Sobell et al. 1988). The LDH takes about 20–30 minutes to complete.

The other two lifetime drinking measures have seen limited use and have each been evaluated in one study. Neither measure has involved clinical populations. The Concordia Lifetime Drinking Questionnaire (CLDQ) assesses lifetime drinking as well as drinking in the 30 days before the last drink (Chaikelson et al. 1994). The CLDQ, whose drinking questions were adapted from Armor and Polich (1982), is administered in a structured interview format and takes about 20 minutes to complete. Like the TLFB, the CLDQ uses visual aids for reconstructing lifetime drinking patterns.

The newest lifetime drinking measure, the Cognitive Lifetime Drinking History (CLDH) (Russell et al. 1997, 1998), "borrows heavily from Skinner's LDH and employs some of the cognitive techniques from the Sobells' Timeline Followback (TLFB) technique" (Lemmens 1998, p. 31s). Before completing the CLDH, respondents use a calendar to note important life events. The CLDH, a computer-administered interview, uses either a floating or a fixed interval (i.e., discrete time periods) and has demonstrated satisfactory reliability for estimates of times intoxicated in a lifetime.

In a thorough review of lifetime drinking measures, Lemmens concluded that while "reliability of lifetime drinking volume varies between 0.90 and 0.67, and is generally quite reliable" (Lemmens 1998, p. 30s), validity measures are

lacking. In another review, Gmel and colleagues (2000) stated that considerable research has been conducted on retrospective lifetime assessments, especially drinking measures, and that reports of distant consumption seem to be as good as and sometimes better than current reports of drinking as a measure of consumption.

Lifetime Drinking Measures

RECOMMENDED USE: To obtain a lifetime or long-term (i.e., greater than the past year) summary of alcohol consumption. These measures take about 20–30 minutes to complete. They provide an overall picture of respondents' alcohol consumption rather than a detailed daily account.

ADVANTAGES: Such measures are advantageous when a longer assessment interval is needed, such as when assessing drinking patterns from adolescence through adulthood, or over a selected time period in the distant past (e.g., natural recovery studies).

LIMITATIONS: Despite reasonably high reliability for an aggregate index of drinking, the LDH lacks precision for the most recent drinking period (Skinner and Allen 1982). Thus, if information about drinking in the past year is needed, a daily drinking estimation procedure should be used.

Quantity-Frequency Measures

QF methods, of which there are many, inquire about "average" or "typical" consumption patterns, usually over a specific time period. These methods, also known as estimation formulas, require respondents to report an average pattern of consumption (e.g., "How many days *on average*—in a specific time interval—did you drink beer, and when you drank beer, *on average* how many beers did you drink?"). Most QF methods repeat these questions for each major alcoholic beverage type (i.e., beer, wine, hard liquor) and then sum across beverage types.

QF measures range from simple scales to sophisticated multidimensional scales. The two major types are single dimensional (e.g., average drinks/day) and multidimensional (e.g., volume variability and volume pattern). The simplest QF measure assesses amount of drinking on average drinking days (Q) and the average number of days when alcohol was consumed (F). The two numbers (i.e., Q and F) are multiplied to derive an estimated total volume referred to as "QF." The multidimensional measures classify individuals into drinker categories based on cross-classifications of quantity and frequency of drinking. The number of drinking categories that have been used for multidimensional measures ranges from 3 to more than 10. For more information, readers are referred to an excellent review of QF methods, including their development, rationale, questionnaire items, and a list of studies (Room 1990).

Although there are several QF variants, in tables 1A and 1B all measures are combined under one category labeled "QF measures." To better understand the variability inherent in QF measures, table 3 lists the major QF measures, the types of drinking data that can be calculated, and the assessment period over which they can collect data.

For all QF measures the following two variables can be calculated: *average quantity per occasion*—average or typical amount of drinking on a given day—and *average frequency per occasion*—how often in a given time interval (e.g., per week, per month) a person consumes the average quantity. Because QF methods ask for average amounts, some methods have included measures of variability or maximum consumption to gather data for occasional heavy drinking. Thus, for some methods *maximum quantity* and *frequency of the maximum quantity* are also obtained. *Variability of quantity per occasion* was introduced in some methods to assess the proportion of drinking occasions in which different numbers of drinks (e.g., 1–2, 5–9, ≥ 10) were consumed.

The first QF measure, developed 50 years ago (Straus and Bacon 1953), classified drinkers by their "typical" drinking patterns. Although this first QF measure inquired about drinking in the past year, subsequent measures have assessed drinking over shorter intervals such as the past 30 days. By today's standards, the first QF measure was primitive because it only asked for the average amount consumed on a given occasion and the average frequency of drinking for three beverage types.

One major criticism of early QF measures was that by only measuring quantity and frequency there was no indication of the variability of a respondent's drinking (Room 1990). Thus, early QF measures were not sensitive to individuals who drank infrequently and consumed large amounts when they drank. For example, consider the following three drinking patterns: someone who drinks 2 drinks every day for a week, someone who drinks 14 drinks on a single day, and someone who has 7 drinks 2 days a week. Although all three patterns result in the same total amount consumed per week (i.e., 14 drinks), if they are extended out over several years they would not only represent very different drinking styles but would also result in different health risks. Recognizing this problem, Cahalan and his colleagues developed two alternative QF measures that assessed the variability of drinking habits (Alanko 1984; Room 1990). For each beverage type, these two methods inquired about the frequency of drinking and the "proportion of drinking occasions" for the various numbers of drinks. The category classifications and calculations for both measures are described in detail elsewhere (Cahalan et al. 1969).

The first measure, Quantity-Frequency Variability (QFV) Index, extended the QF measure by measuring maximum quantity per occasion (Cahalan et al. 1969). The proportion of occasions for the QFV Index is determined by asking how often the person consumed 5+, 3–4, and 1–2 drinks. Proportions are defined on a 4-point scale ranging from nearly every time

TABLE 3.—Summary of quantity-frequency drinking measures

Measure (reference)	Average/ typical quantity per occasion	Average frequency per occasion	Variability of quantity per occasion	Maximum quantity	Frequency of maximum quantity	Aggregate volume[1]	Assessment timeframe
Quantity-Frequency (Straus and Bacon 1953)	•	•					Past year
Volume-Variability Index (Cahalan and Cisin 1968)	•	•	•			•	Average/month
Quantity-Frequency Variability Index (Cahalan et al. 1969)	•	•					Average/month
Volume-Pattern Index (Bowman et al. 1975)	•[2]	•[2]				•	Maximum of 3 months
NIAAA Quantity Frequency (Armor et al. 1978)	•	•					Past 30 days
Khavari Alcohol Test (Khavari and Farber 1978)[2]	•	•			•		None stated
Composite Quantity Frequency Index (Polich and Orvis 1979)	•	•					30 days before last drink for quantity-frequency, past year for high frequency
Rand Quantity Frequency (Polich et al. 1981)	•	•	•				30 days before last drink
Graduated-Frequency Measure (Clark and Midanik 1982; Midanik 1994)[3]	•	•	•	•	•	•	Past 12 months
Lifetime Drinking History (Skinner and Sheu 1982)	•	•	•	•	•[4]	•	Lifetime
Concordia Lifetime Drinking Questionnaire (Chaikelson et al. 1994)	•	•		•[5]			Lifetime/30 days before last drink
Cognitive Life Drinking History (Russell et al. 1997)	•	•	•	•	•	•	Lifetime

[1] Average drinks per day in the interval.

[2] Modified version of Quantity-Frequency Variability Index (Cahalan et al. 1969).

[3] Combined beverage use (e.g., two beers and one glass of wine).

[4] Frequency of maximum amount category added by L.C. Sobell et al. (1988).

[5] Current drinking questions from Armor and Polich (1982).

to never. Based on respondents' answers regarding the alcoholic beverage consumed most often, a complicated classification schema with 11 classes of quantity and variability components was developed (Cahalan et al. 1969). The QFV Index is derived by combining the quantity-variability classification for the beverage most often consumed with frequency of drinking any alcoholic beverage. Lastly, although somewhat arbitrary, these QFV classifications led to the creation of five drinker groups: heavy, moderate, light, infrequent, and abstainers.

The second QF variability measure, the Volume-Variability (VV) Index, classifies drinking into eight categories (see Cahalan et al. 1969, p. 215) based on the aggregate volume (Q x F) and the maximum quantity variables (Cahalan and Cisin 1968). The VV Index was developed based on the "principle that spacing or bunching of drinks is more important than aggregate volume in characterizing an individual's drinking patterns" (Cahalan et al. 1969, p. 17). The VV Index computes a person's average daily volume (multiplying the frequency of drinking each beverage—i.e., number of days drinking per 30 days—by estimated quantity of the beverage consumed per occasion) and then classifies drinkers as to whether they ever had as many as 5 drinks on one occasion (Cahalan et al. 1969).

Cahalan and his colleagues recommended using the VV Index because it has "all of the useful characteristics of the QFV Index and also preserves the distinction between those who consume a given volume by bunching or massing their drinks and those who space them out" (Cahalan et al. 1969, p. 17). Compared with the QFV Index, the VV Index is more sensitive to differences in the middle range of drinking (noted in Khavari and Farber 1978). As additional surveys were conducted, it became apparent that the upper range category of 5+ drinks was insensitive to very heavy drinking (i.e., substantial numbers of individuals drink at these levels).

Consequently, Cahalan and his colleagues combined two methods: "proportion of occasions" questions for 5+, 3–4, and 1–2 drinks and nonbeverage-specific questions for 8–11 and 12+ drinks for a 1-year reporting period (Room 1990).

The Khavari Alcohol Test (Khavari and Farber 1978), a 12-question version of the QF method used by Cahalan and his colleagues (1969), asks respondents to report their usual frequency of drinking, the usual amount consumed per occasion, the maximum amount consumed on any one occasion, and the frequency of the maximum amount. These four questions are repeated for each of three alcoholic beverage types. Respondents are first categorized into 1 of 11 frequency categories, and then their drinking is plotted and compared with normative values.

In an effort to avoid the classification of drinkers into discrete categories, Bowman and his colleagues (1975) developed a continuous measure reflecting the volume and pattern of a person's drinking. The volume component is an aggregate volume measure derived from QF data, and the pattern component is an adjusted standard deviation measure indicating the degree of volume variability over time. Although the Volume-Pattern Index was an attempt to improve on previous QF methods, it has been criticized as cumbersome in terms of data manipulation and transformations (Khavari and Farber 1978). Further, because it asks for very detailed drinking information, it can take 30–60 minutes to complete.

The NIAAA QF measure, a variant of the original QF measure, was used in national drinking surveys conducted in the early 1970s as part of NIAAA's public service advertisement campaigns. NIAAA also used this QF measure in its evaluation of alcohol treatment centers (Armor et al. 1978). The Rand QF (Polich et al. 1981), like the NIAAA QF, asks respondents to recall how much they consumed on a typical day during the 30 days before their last drink for each beverage type. Respondents are also asked to recall the

number of days drinking at or exceeding fairly high levels (i.e., 6–9 drinks, 10+ drinks) during this same interval. The intent of the Rand QF is to determine a person's typical drinking pattern and then to assess atypical, heavy drinking.

The Composite QF Index (Polich and Orvis 1979), an unusual QF hybrid, asks about the 30 days before the last drinking occasion for all alcoholic beverages combined (versus specific types of alcohol). It also asks about the frequency of heavy drinking (i.e., 8+ drinks) in the past year. By adding questions for the past year to the typical 30-day window, this measure assesses recent and distant heavy drinking.

The LDH (Skinner and Sheu 1982) and related lifetime drinking measures are specialized QF methods that were described earlier. Unlike other QF measures, these measures ask about lifetime drinking.

The Graduated-Frequency (GF) Measure (Clark and Midanik 1982; Midanik 1994) was developed in response to criticisms that QF measures failed to account for occasions when different types of beverages were combined (e.g., beer and whiskey on the same day). The GF Measure asks respondents to report the frequency of their drinking for different levels of drinking (e.g., 1–2 drinks or 3–4 drinks; highest level is most ever consumed) in the last year for combined beverage types. The GF and LDH methods are among the few QF measures that ask questions for all alcoholic beverages combined. Because there are no standardized ways to assess alcohol consumption in epidemiologic studies, one study compared three widely used methods (QF, GF, and weekly drinking recall) for estimates of high-risk drinking and consequences (Rehm et al. 1999). The GF Measure yielded much higher estimates of the prevalence of high-risk drinking and consequences.

Quantity-Frequency (QF) Measures

RECOMMENDED USE: QF methods generally provide reliable information about total consumption (quantity) and number (frequency) of drinking days. They are most useful when a quick measure of drinking is needed and when drinking is unpatterned.

ADVANTAGES: QF methods provide a quick and easy estimate when information needs are limited to a rough estimate of the total amount consumed or of the total number of drinking days in an interval, or if time is at a premium (e.g., physician's office) and knowledge of atypical drinking is not needed.

LIMITATIONS: There is no shortage of reviews and critiques of problems with QF methods (Polich and Kaelber 1985; Room 1990; L.C. Sobell and Sobell 1992). Although the GF Measure escapes many of the limitations that befall other QF methods, it is at the expense of a much longer administration time. QF measures reflect less drinking, and they tend to misclassify drinkers compared with daily diary or TLFB reports. Many QF methods also do not ask for different types of alcoholic beverages consumed (e.g., three beers and two glasses of wine) on the same day. Unfortunately, when QF methods (e.g., the Volume-Pattern Index and the GF Measure) do ask about combined beverage use, the result is a longer administration time. In addition, QF methods cannot provide a picture of unpatterned fluctuations in drinking. Finally, because days of sporadic heavy drinking commonly and frequently occur in clinical populations, assessment of such drinking is important. Unfortunately, with the exception of the GF Measure, such drinking days are not captured by QF methods.

COMPARISONS AMONG DRINKING MEASURES

Room (1990) reported that when two different studies added questions on the frequency of consuming 8+ drinks as compared with a cutoff with 5+ drinks, the total average drinking volume was raised by 16 percent and 36 percent, respectively. This should not be surprising given the early criticisms of QF methods as insensitive to atypical heavy drinking days. More recently, Midanik (1994) compared a typical QF measure with the GF Measure. The latter measure involved a series of questions about single and combined beverage use that yielded measures of the frequency of consuming specific numbers of drinks over the past year. Overall, the GF Measure yielded higher estimates of alcohol use, while the QF measure provided a higher estimate of lighter drinkers and a lower estimate of heavier drinkers.

As noted earlier (Kuhlhorn and Leifman 1993), a report describing two Swedish alcohol surveys showed significant differences in their coverage of beverage sales reports, with a daily drinking format yielding considerably greater coverage (75 percent) of beverage sales compared with a QF method (28 percent).

Rehm and his colleagues compared three ways of assessing high-risk drinking in surveys—GF, typical QF, and weekly drinking (i.e., 7 days before the survey)—and found that "the GF measure had much higher sensitivity than the other measures for identifying potentially harmful levels of consumption . . . because it is more effective in capturing episodes of very high consumption" (Rehm et al. 1999, p. 222). While they also concluded that a brief QF measure would be sufficient if a genuine average across all drinking situations was the desired effect, for many cultures and social groups the GF Measure would be preferred.

Use of varying recall strategies resulted in twice as many older adults being classified as nondrinkers by short interval measures (i.e., 7-day daily diary and 7-day QF) compared with a longer interval (Werch 1989). This finding highlights the problem of using a short timeframe to gather data for infrequent drinkers. The 7-day retrospective diary also resulted in greater reported daily alcohol use and a greater number of drinks reported consumed per week than either the 7-day or 28-day QF measure. Further, the GF Measure, because of its beverage-specific assessment, has been shown to result in higher drinking estimates than typical QF measures. The GF Measure captures days of sporadic heavy drinking better than QF measures because of the former's elaborate series of questions. A study comparing three QF methods—global, beverage specific, and beverage specific with drink size—found that adding beverage type and drink size estimates to QF measures increased reported daily alcohol consumption (Williams et al. 1994).

Several studies have compared various QF measures with the TLFB or similar daily drinking measures and have found that daily measures almost always provide greater estimates of drinking than QF measures (Cooney et al. 1984; M.B. Sobell et al. 1986; Fitzgerald and Mulford 1987; Redman et al. 1987; L.C. Sobell et al. 1988; Werch 1989; Flegal 1990; Saunders and Conigrave 1990; O'Hare et al. 1991; Duffy and Alanko 1992; Lemmens et al. 1992). Because studies comparing daily drinking measures and QF measures have been reviewed in considerable detail elsewhere (see L.C. Sobell and Sobell 1992), they will not be reviewed here except for a few notable findings.

Two studies that compared data from the TLFB and different QF measures found large differences between reports on the TLFB compared with QF drinker classifications (M.B. Sobell et al. 1986; L.C. Sobell et al. 1988; L.C.

Sobell and Sobell 1992). For example, one QF method that classified drinkers as heavy consumers found that their TLFB reports for amount consumed over 90 days ranged from 30 to 370 standard drinks. Similar wide-ranging classifications occurred for the variables mean drinks per drinking day and number of days drinking. Other studies have found similar discrepancies. For example, in a study assessing dietary consumption where drinking was recorded as part of a QF dietary questionnaire or a self-reported diet diary (i.e., no separate alcohol data collection), 31 percent of heavy drinkers identified by their daily diary reports were classified as moderate drinkers by QF methods (Flegal 1990). In another study, the QF methods failed to detect 78 percent of heavy drinkers identified by daily diary reports (Redman et al. 1987).

One study more than others illustrates the problem of QF methods' insensitivity for assessing atypical drinking (Fitzgerald and Mulford 1987). After asking a routine set of QF questions, seven additional questions were asked inquiring about *atypical* drinking. As a result of these questions, 35 percent of all adults reported more drinking. Moreover, "the addition of atypical drinking to ordinary consumption increased the total consumption estimate for adults by 14 percent" (Fitzgerald and Mulford 1987, p. 208). Interestingly, the GF Measure (Hilton 1989) and a recent occasions recall measure (Wyllie et al. 1994) both showed consistent results with a daily diary (30 and 7 days, respectively) when data were examined at a population level.

Although daily drinking measures are typically superior to QF measures, a recent study (L.C. Sobell et al. in press) found good correspondence between a QF and a TLFB measure. As part of a large ($N = 825$) community self-help intervention (L.C. Sobell et al. 1996, 2002), drinking was assessed in two ways: mailed-in 360-day TLFB assessment and telephone Quick Drinking Screen (QDS) (QF summary measure). Five measures of

consumption comprising the QDS were found to yield very similar data (e.g., days drinking ≥ 5 drinks/day in the past year: TLFB = 164.4, QDS = 176.5; drinks per week past year: TLFB = 31.9, QDS = 31.3). Although the QDS has an advantage in terms of speed and brevity, like all QF summary measures it does not allow for an evaluation of temporal patterning or variability of drinking.

The QDS, besides being used for screening, was also used to collect followup data for alcohol abusers who were not willing to complete a lengthy followup interview by mail or phone. This resulted in an additional 29 percent (189/656) of respondents providing drinking data at the 1-year followup (L.C. Sobell et al. 2002). A brief variant of Form 90 has similarly been used to gather data for clients unwilling or unable to complete a followup interview (Miller and Del Boca 1994).

A problem shared by retrospective measures, whether they are daily drinking or QF measures, is forgetting. This is exemplified in studies that have compared retrospective measures, such as the TLFB with the concurrent measure of self-monitoring. Even though both methods measure daily drinking, studies have found that self-monitoring resulted in a slightly higher frequency of drinking days compared with TLFB or daily diary methods (Samo et al. 1989; M.B. Sobell et al. 1989; Lemmens et al. 1992), but no differences in reported quantity per drinking day. This suggests that errors are mainly related to forgetting rather than minimization of drinking. Research indicates that errors in judgments for the frequency of other behaviors relates to memory and contextual cues (Menon and Yorkston 2000).

Another study (Searles et al. 2000) compared drinking reports using an interactive voice response (IVR) system with the TLFB. Using an IVR system, people call a toll-free number daily and respond to telephone prompts to report their drinking for the previous day. While correlations between the IVR and TLFB for amount consumed, drinking days, and heavy drinking

days were modest, there was large variability in individual participant correlations between their TLFB and IVR reports. This replicates a finding by Vuchinich et al. (1985), who found strong correlations between TLFB aggregate data (e.g., total number of days drinking) but found lower correspondence for day-by-day reports. This suggests that precise day-by-day reports obtained at two different times or by two different methods are inconsistent but that overall reported levels of consumption are reliable.

More research is needed on the IVR procedure: (a) it has not been evaluated with alcohol abusers; (b) it has not been evaluated in a clinical setting; (c) there has been no validation that respondents have been alcohol free when providing IVR reports; and (d) there has been no demonstration that IVR produces reports that are superior to self-monitoring, a much less costly alternative concurrent measure. In addition, concerns about reactivity with this procedure are similar to those for daily self-monitoring logs. That is, the very act of reporting one's drinking may affect an individual's drinking, and concurrent reporting methods might make it difficult to identify treatment effects in some situations (e.g., controlled trials). Another problem with the IVR procedure is that it is unknown what level of compliance would occur without incentives. Searles et al. (2000) paid participants 50¢ per day for reporting, plus a bonus of $1 per week for reporting all 7 days, and a bonus of $500 for participation in the 2-year study. All participants also competed for entry into a drawing for a $6,000 prize, to be divided among those with the best calling records ($3,000 for the best record). Participants were also paid $25 for their interviews every 3 months. Interestingly, even with incentives, Searles et al. (2000) reported that a third of participants refused to continue when the initial 7-month study was extended to 24 months.

A final and important issue regarding concurrent versus retrospective measures is that concurrent measures have little utility for assessment of pretreatment drinking. The only way that pretreatment data can be gathered prospectively is to have individuals self-monitor before they begin treatment. Such a procedure has two serious drawbacks. First, it would necessitate delaying treatment for the sole purpose of gathering pretreatment data prospectively, and such a procedure seems ethically objectionable. Second, the self-monitoring might be reactive, raising questions about whether the assessment data are representative of pretreatment drinking. Consequently, retrospective methods are likely to be the procedure of choice for gathering pretreatment assessment information.

In summary, there are two main dimensions along which self-reported measures of alcohol consumption differ: (a) *summary* (e.g., QF) versus *daily drinking* measures (e.g., TLFB) and (b) *retrospective* (e.g., TLFB and QF) versus *concurrent* (e.g., self-monitoring and IVR) measures. In terms of summary versus daily drinking measures, although QF measures can provide reliable information about total consumption and number of drinking days, with the exception of the GF Measure they have some serious limitations when compared with daily recall methods:

- They do not measure sporadic heavy drinking, which is clinically important.
- Many QF methods do not correct for days when more than one type of alcoholic beverage is consumed.
- QF methods cannot provide a temporal picture of drinking patterns.
- Newer variants of QF methods, while designed to more accurately reflect actual drinking, take more time to collect drinking data, thus negating the advantage of brevity of early QF methods.

In terms of retrospective versus concurrent measures, it is recommended that a daily drinking estimation procedure be used to gather pre- and posttreatment information for clinical and research purposes. For within-treatment data, self-monitoring can be used. The downside of using retrospective measures to gather pretreatment data

and concurrent measures to gather followup data would be the introduction of a methodological bias that works against finding treatment effects (i.e., even if there were no treatment effect, one would expect retrospective reports of pretreatment drinking to be lower than prospective reports of posttreatment drinking). Thus, it may be better to use retrospective measures for both purposes, an approach that would be expected to keep errors consistent across temporal intervals. Ultimately, the choice of what measure to use will depend on its intended purpose (Leigh 2000).

DEVELOPING A CONSENSUS

In April 2000, 40 researchers from 12 countries came together at a thematic conference of the Kettil Bruun Society for Social and Epidemiological Research on Alcohol (Dawson and Room 2000). The conference had three goals, one of which was "developing a consensus set of questionnaire items for measuring alcohol consumption, including both a minimum set of essential items for addressing policy concerns and other desirable items for more extensive research purposes" (Dawson and Room 2000, p. 2). This ambitious goal resulted in several recommendations (e.g., temporal reference period for assessing drinking; quantity thresholds) that collectively are a major step forward in developing consensus on what has always been a thorny issue—when and how to best measure alcohol use. Although it is clear from the recommendations that there is no flawless measure and that the best measure will depend on the purpose of the assessment, the recommendations are important and have been summarized in the appendix to this chapter. Readers interested in the rationale and discussion surrounding these recommendations are referred to the source article (Dawson and Room 2000) and 12 other articles that were part of a special issue on measuring alcohol consumption in the *Journal of Substance Abuse* (Volume 12, 2000, pp. 1–212).

SUMMARY

Since the first QF method appeared half a century ago, the assessment of drinking has advanced considerably. Today a variety of measures are available to retrospectively assess drinking over varying time intervals. Many of these measures have both clinical and research utility with a variety of drinker groups. Although several studies suggest that memory aids can be used to enhance recall of drinking (Midanik and Hines 1991; L.C. Sobell and Sobell 1992; Hammersley 1994; Single and Wortley 1994), additional research evaluating contextual cues to improve recall accuracy is encouraged. It is important to remember that almost all drinking measures are retrospective and, as such, they require people to provide their "best estimate" of their past drinking. Thus, some amount of error is expected.

Two articles comparing different ways of measuring risky or hazardous drinking in surveys end with the same recommendations as this chapter. In the first article, Rehm and his colleagues (1999) compared three ways of assessing high-risk drinking and concluded that we still have much to learn about how best to assess alcohol consumption and that the method used should be determined by the objective of the assessment. In the second article, Dawson concluded that efforts to promote the use of a "single 'best' measure of *any* aspect of alcohol consumption may be unrealistic or even counterproductive, simply because the measures that work best for one application may not be the best for *all* applications" (Dawson 2000, p. 91).

Finally, consistent with the intent of this volume and as recognized by others (L.C. Sobell et al. 1994; Treatment Improvement Protocol Series 35 Consensus Panel 1999), drinking measures, like other alcohol assessment measures, should be designed whenever possible to have research and clinical utility.

APPENDIX: DRINKING GUIDELINES[1]

Recommendations: For developing a consensus set of questionnaire items for measuring alcohol consumption, including both a minimum set of essential items for addressing policy concerns and other desirable items for more extensive research purposes.

2.1 *Reference period for reporting*

a. A past-year reference period is recommended for linking alcohol consumption with alcohol-related consequences.

b. To characterize drinking occasions at the individual level, a period of varying length that incorporates the past four drinking occasions is recommended.

c. To characterize drinking occasions at the aggregate level, asking about consumption on the last one or two occasions might be considered, though this approach is not satisfactory for characterizing the individual respondent's drinking.

2.2 *Measuring frequency of drinking*

a. Questions on drinking frequency should not be asked in a totally open-ended format (e.g., number of times per year).

b. Frequency should be asked in terms of pre-specified frequency range categories or in terms of times per week, falling back on times per month or per year for infrequent drinkers.

c. Frequency categories should be arrayed in terms of descending order; i.e., the most frequent first.

2.3 *Measuring quantity of drinks: per occasion or per day?*

a. For maximum cross-cultural comparability, quantities should be asked in terms of number of drinks per day, with a day defined to include continued drinking past midnight.

2.4 *Asking specified quantities "up" or "down"?*

a. Additional methodological studies are recommended to determine whether it is preferable to ask about specific quantity ranges in ascending or descending order.

2.5 *Quantity thresholds*

a. Quantity thresholds should, at minimum, include numbers of standard drinks corresponding to 144 g, 96 g, and 60 g ethanol. Additional lower quantity thresholds are desirable if the questions are used to estimate volume.

2.6 *Different thresholds for women and men?*

a. In view of the continuing debate concerning different quantity thresholds for men and women, a prudent approach is to select a single set of quantity thresholds or bands that include all the cut points thought to represent hazardous and/or harmful consumption for both men and women, and to confirm gender differences in the course of analysis, rather than by building assumptions into the questions used to obtain the data.

2.7 *Cumulative or discrete quantity bands in "graduated frequency" approaches?*

a. Cumulative quantity bands, beginning with the larger quantity thresholds first and working down, are recommended for asking about the frequency of drinking different amounts in instruments intended for cross-cultural use.

2.8 *Usual-quantity questions*

a. A single question on usual quantity should not form the sole basis for estimating volume of consumption, but it is useful to ask for comparative purposes.

2.9 *Specific beverage types*

a. Questions on individual beverage types should be included. If space does not permit asking detailed questions on quantity and frequency for each beverage type, limited questions on frequency of drinking each beverage or type of beverage most frequently consumed are still useful.
b. The types of beverages included must vary to reflect individual countries' consumption patterns.

2.10 *More precise measurement of indicators of attained BALs*

a. Questions on duration of drinking occasions and body mass index (height, weight, age, gender) should be included to interpret effects of quantity consumed on BALs.

2.11 *Context-of-drinking questions*

a. Recommended measures of drinking context include at meals vs. not at meals, weekday vs. weekend, in public vs. at home, alone vs. others.

2.12 *Frequency of getting drunk/intoxication*

a. Questions on frequency of drunkenness/ intoxication are preferable to those on feeling the effects.
b. Although variable in their own right, these should not be used as proxies for frequency of heavy drinking.

2.13 *Minimum set of questions on drinking amount and pattern*

a. abstention—lifetime and past 12 months
b. overall frequency of drinking (all alcoholic beverages together)
c. usual quantity of drinking (all alcoholic beverages together)
d. frequency of consuming > 60 g ethanol in a day (1st alternative: if usual quantity was > 60g, ask frequency of consuming > 96 g; alternative: largest amount drunk in a day in the past 12 months and how often that amount was consumed)
e. frequency of drunkenness (if possible)

2.14 *Recommended set of questions on drinking amount and pattern*

a. abstention—lifetime and past 12 months
b. largest amount drunk in last 12 months (maximum quantity), all beverages together
c. graduated frequencies questions, all beverages together:
   cut-offs: * ≈ 24 and/or ≈ 36, 60, 96, 144, 240g for largest amount (less desirable alternative: frequency of drinking > 60 g)
d. overall frequency of drinking, all beverages together
   (critical if graduated frequencies questions cannot be summed to estimate overall frequency of drinking, e.g., if only asking frequency of drinking > 60 g; desirable even when graduated frequencies are asked)
e. beverage-specific frequencies of drinking
   (if there is an emphasis on measuring volume of drinking, frequency categories should be fairly fine, e.g.: twice a day, daily, 5–6 times a week/nearly every day, 3–4 times a week, once or twice a week, 2–3 times a month, once a month, 6–11 times a year, 1–5 times a year)
f. beverage-specific usual quantities of drinking
g. beverage-specific size of usual drink

h. frequency of drunkenness and number of drinks to feel drunk
- usual quantity of drinking, all beverages combined
- frequency of consuming maximum quantity, all beverages combined (high priority if graduated frequencies questions are not asked)
- frequency of drinking "enough to feel the effects" and number of drinks for that
- beverage-specific maximum quantities and associated frequencies
- body weight and height
- context of drinking and duration of drinking occasions

3. *Aggregating drinking patterns for analysis*

a. Volume of drinking
   Frequency of 5+ or frequency of 8+ or maximum Q
b. Volume of drinking
   Variance in volume or volume-specific binge measure (higher quantity cut-off for higher volumes)
c. Frequency of drinking
   Usual/average quantity per occasion
   Variance of quantity or frequency of 5+, etc.

REFERENCES

Alanko, T. An overview of techniques and problems in the measurement of alcohol consumption. In: Smart, R.G.; Cappell, H.; Glaser, F.; Israel, Y.; Kalant, H.; Popham, R.E.; Schmidt, W.; and Sellers, E.M., eds. *Research Advances in Alcohol and Drug Problems*. Vol. 8. New York: Plenum Press, 1984. pp. 209–226.

American Psychiatric Association. *Handbook of Psychiatric Measures*. Washington, DC: the Association, 2000.

Armor, D.J., and Polich, M.J. Measurement of alcohol consumption. In: Pattison, E.M., and Kaufman, E., eds. *Encyclopedic Handbook of Alcoholism*. New York: Gardner Press, 1982. pp. 72–80.

Armor, D.J.; Polich, J.M.; and Stambul, H.B. *Alcoholism and Treatment*. New York: John Wiley & Sons, 1978.

Babor, T.F. Reliability of the ethanol dependence syndrome scale. *Psychol Addict Behav* 10(2):97–103, 1996.

Babor, T.F.; Brown, J.; and Del Boca, F.K. Validity of self-reports in applied research on addictive behaviors: Fact or fiction? *Addict Behav* 12:5–32, 1990.

Babor, T.F.; Steinberg, K.; Anton, R.; and Del Boca, F. Talk is cheap: Measuring drinking outcomes in clinical trials. *J Stud Alcohol* 61(1):55–63, 2000.

Bowman, R.S.; Stien, L.I.; and Newton, J.R. Measurement and interpretation of drinking behavior. *J Stud Alcohol* 36:1154–1172, 1975.

Brown, J.; Kranzler, H.R.; and Del Boca, F.K. Self-reports by alcohol and drug abuse inpatients: Factors affecting reliability and validity. *Br J Addict* 87:1013–1024, 1992.

Cahalan, D., and Cisin, I.H. American drinking practices: Summary of findings from a national probability sample. *Q J Stud Alcohol* 29:130–151, 1968.

Cahalan, D.; Cisin, I.H.; and Crossley, H.M. *American Drinking Practices*. New Brunswick, NJ: Rutgers Center of Alcohol Studies, 1969.

Chaikelson, J.S.; Arbuckle, T.Y.; Lapidus, S.; and Gold, D.P. Measurement of lifetime alcohol consumption. *J Stud Alcohol* 55:133–140, 1994.

Clark, W. B., and Midanik, L.T. *Alcohol Use and Alcohol Problems Among U.S. Adults: Results of the 1979 Survey*. Alcohol and Health Monograph No. 1. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism, 1982.

Cooney, N.L.; Baker, L.H.; Kaplan, R.F.; and Meyer, R. Measurement of recent drinking history and alcohol dependence: A validity study. Paper presented at the Third International Conference on Treatment of Addictive Behaviors, Scotland, August 1984.

Cunningham, J.A.; Ansara, D.; Wild, T.C.; Toneatto, T.; and Koski-Jännes, A. What is the price of perfection? The hidden costs of using detailed assessment instruments to measure alcohol consumption. *J Stud Alcohol* 60(6): 756–758, 1999.

Dawson, D.A. Alternative measures and models of hazardous consumption. *J Subst Abuse* 12(1-2):79–91, 2000.

Dawson, D.A., and Room, R. Towards agreement on ways to measure and report drinking patterns and alcohol-related problems in adult general population surveys: The Skarpo Conference overview. *J Subst Abuse* 12(1-2): 1–21, 2000.

Del Boca, F.K., and Noll, J.A. Truth or consequences: The validity of self-report data in health services research on addictions. *Addiction* 95:S347–S360, 2000.

Drake, R.E.; McHugo, G.J.; and Biesanz, J.C. The test-retest reliability of standardized instruments among homeless persons with substance use disorders. *J Stud Alcohol* 56(2):161–167, 1995.

Duffy, J.C., and Alanko, T. Self–reported consumption measures in sample surveys: A simulation study of alcohol consumption. *J Off Stat* 8:327–350, 1992.

Fitzgerald, J.L., and Mulford, H.A. Self-report validity issues. *J Stud Alcohol* 48:207–211, 1987.

Flegal, K.M. Agreement between two dietary methods in the measurement of alcohol consumption. *J Stud Alcohol* 51:408–414, 1990.

Gmel, G.; Rehm, J.; Room, R.; and Greenfield, T.K. Dimensions of alcohol-related social and health consequences in survey research. *J Subst Abuse* 12(1-2):113–138, 2000.

Göransson, M., and Hanson, B.S. How much can data on days with heavy drinking decrease the underestimation of true alcohol consumption? *J Stud Alcohol* 55(6):695–700, 1994.

Grant, K.A.; Tonigan, J.S.; and Miller, W.R. Comparison of three alcohol consumption measures: A concurrent validity study. *J Stud Alcohol* 56:168–172, 1995.

Hammersley, R. A digest of memory phenomena for addiction research. *Addiction* 89:283–293, 1994.

Harris, K.B., and Miller, W.R. Behavioral self-control training for problem drinkers: Components of efficacy. *Psychol Addict Behav* 4:82–90, 1990.

Hilton, M.E. Inconsistent responses to questions about alcohol consumption in specified settings. *Am J Drug Alcohol Abuse* 12: 403–413, 1986.

Hilton, M.E. A comparison of a prospective diary and two summary recall techniques for recording alcohol consumption. *Br J Addict* 84: 1085–1092, 1989.

Kavanagh, D.J.; Sitharthan, T.; Spilsbury, G.; and Vignaendra, S. An evaluation of brief correspondence programs for problem drinkers. *Behav Ther* 30(4):641–656, 1999.

Khavari, K.A., and Farber, P.D. A profile instrument for the quantification and assessment of alcohol consumption: The Khavari Alcohol Test. *J Stud Alcohol* 39:1525–1539, 1978.

Korotitsch, W.J., and Nelson-Gray, R.O. An overview of self-monitoring research in assessment and treatment. *Psychol Assess* 11(4):415–425, 1999.

Kuhlhorn, E., and Leifman, H. Alcohol surveys with high and low coverage rate: A comparative analysis of survey strategies in the alcohol field. *J Stud Alcohol* 54:542–554, 1993.

Leigh, B.C. Using daily reports to measure drinking and drinking patterns. *J Subst Abuse* 12(1-2): 51–65, 2000.

Lemmens, P.H. Measuring lifetime drinking histories. *Alcohol Clin Exp Res* 22 (Suppl. 2): 29s–36s, 1998.

Lemmens, P.H., and Knibbe, R.A. Seasonal variation in survey and sales estimates of alcohol consumption. *J Stud Alcohol* 54:157–163, 1993.

Lemmens, P.; Tan, E.S.; and Knibbe, R.A. Measuring quantity and frequency of drinking in a general population survey: A comparison of 5 indices. *J Stud Alcohol* 53:476–486, 1992.

Maisto, S.A.; McKay, J.R.; and Connors, G.J. Self-report issues in substance abuse: State of the art and future directions. *Behav Assess* 12:117–134, 1990.

Mäkelä, K. *Measuring the Consumption of Alcohol in the 1968-69 Alcohol Consumption Study*. Report No. 2. Helsinki, Finland: Social Research Institute of Studies, 1971.

Menon, G., and Yorkston, E.A. The use of memory and contextual cues in the formation of behavioral frequency judgments. In: Stone, A.A.; Turkkan, J.S.; Bachrach, C.A.; Jobe, J.B.; Kurtzman, H.S.; and Cain, V.S., eds. *The Science of Self-Report: Implications for Research and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000. pp. 63–79.

Midanik, L. The validity of self-reported alcohol consumption and alcohol problems: A literature review. *Br J Addict* 77:357–382, 1982.

Midanik, L.T. Comparing usual quantity/frequency and graduated frequency scales to assess yearly alcohol consumption: Results from the 1990 United States National Alcohol Survey. *Addiction* 89:407–412, 1994.

Midanik, L.T., and Hines, A.M. "Unstandard" ways of answering standard questions: Protocol analysis in alcohol survey research. *Drug Alcohol Depend* 27:245–252, 1991.

Miller, W.R., and Del Boca, F.K. Measurement of drinking behavior using the Form 90 family of instruments. *J Stud Alcohol* (Suppl. 12): 112–118, 1994.

Miller, W.R., and Marlatt, G.A. *Manual for the Comprehensive Drinker Profile*. Odessa, FL: Psychological Assessment Resources, 1984.

Nelson, R.O., and Hayes, S.C. Theoretical explanations for reactivity in self-monitoring. *Behav Modif* 5:3–14, 1981.

O'Hare, T. Measuring alcohol consumption: A comparison of the retrospective diary and the quantity-frequency methods in a college drinking survey. *J Stud Alcohol* 52:500–502, 1991.

O'Hare, T.; Bennett, P.; and Leduc, D. Reliability of self-reports of alcohol use by community clients. *Hosp Community Psychiatry* 42:406–408, 1991.

Poikolainen, K., and Kärkkäinen, P. Nature of questionnaire options affects estimates of alcohol intake. *J Stud Alcohol* 46:219–222, 1985.

Polich, J.M., and Kaelber, C.T. Sample surveys and the epidemiology of alcoholism. In: Schuckit, M.A., ed. *Alcohol Patterns and Problems*. New Brunswick, NJ: Rutgers Center of Alcohol Studies, 1985. pp. 43–77.

Polich, J.M., and Orvis, B.R. *Alcohol Problems: Patterns and Prevalence in the U. S. Air Force*. Santa Monica, CA: Rand, 1979.

Polich, J.M.; Armor, D.J.; and Braiker, H.B. *The Course of Alcoholism: Four Years After Treatment*. New York: John Wiley & Sons, 1981.

Project MATCH Research Group. Project MATCH: Rationale and methods for a multisite clinical trial matching patients to alcoholism treatment. *Alcohol Clin Exp Res* 17(6):1130–1145, 1993.

Redman, S.; Sanson-Fisher, R.W.; Wilkinson, C.; Fahey, P.P.; and Gibberd, R.W. Agreement between two measures of alcohol consumption. *J Stud Alcohol* 48:104–108, 1987.

Rehm, J.; Greenfield, T.K.; Walsh, G.; Xie, X.; Robson, L.; and Single, E. Assessment methods for alcohol consumption, prevalence of high risk drinking and harm: A sensitivity analysis. *Int J Epidemiol* 28(2):219–224, 1999.

Room, R. Measuring alcohol consumption in the United States: Methods and rationales. In: Kozlowski, L.T.; Annis, H.M.; Cappell, H.D.; Glaser, F.B.; Goodstadt, M.S.; Israel, Y.; Kalant, H.; Sellers, E.M.; and Vingilis, E.R., eds. *Research Advances in Alcohol and Drug Problems*. Vol. 10. New York: Plenum Press, 1990. pp. 39–80.

Russell, M.; Marshall, J.R.; Trevisan, M.; Freudenheim, J.L.; Chan, A.W.; Markovic, N.; Vana, J.E.; and Priore, R.L. Test-retest reliability of the Cognitive Lifetime Drinking History. *Am J Epidemiol* 146(11):975–981, 1997.

Russell, M.; Peirce, R.S.; Vana, J.E.; Nochajski, T.H.; Carosella, A.M.; Muti, P.; Freudenheim, J.; and Trevisan, M. Relations among alcohol consumption measures derived from the Cognitive Lifetime Drinking History. *Drug Alcohol Rev* 17(4):377–387, 1998.

Samo, J.A.; Tucker, J.A.; and Vuchinich, R.E. Agreement between self-monitoring, recall, and collateral observation measures of alcohol consumption in older adults. *Behav Assess* 11:391–409, 1989.

Sanchez-Craig, M., and Annis, H.M. "Self-monitoring" and "recall" measures of alcohol consumption: Convergent validity with biochemical indices of liver function. *Br J Alcohol Alcoholism* 17:117–121, 1982.

Saunders, J.B., and Conigrave, K.M. Early identification of alcohol problems. *CMAJ* 143:1060–1069, 1990.

Schwarz, N. Self-reports: How the questions shape the answers. *Am Psychol* 54(2):93–105, 1999.

Searles, J.S.; Helzer, J.E.; and Walter, D.E. Comparison of drinking patterns measured by daily reports and Timeline Followback. *Psychol Addict Behav* 14(3):277–286, 2000.

Single, E., and Wortley, S. A comparison of alternative measures of alcohol consumption in the Canadian National Survey of Alcohol and Drug Use. *Addiction* 89:395–399, 1994.

Skinner, H.A., and Allen, B.A. Alcohol dependence syndrome: Measurement and validation. *J Abnorm Psychol* 91:199–209, 1982.

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., and Sobell, M.B. Self-report issues in alcohol abuse: State of the art and future directions. *Behav Assess* 12:91–106, 1990.

Sobell, L.C., and Sobell, M.B. Timeline Followback: A technique for assessing self-reported alcohol consumption. In: Litten, R.Z., and Allen, J.P., eds. *Measuring Alcohol Consumption: Psychosocial and Biological Methods*. Totowa, NJ: Humana Press, 1992. pp. 41–72.

Sobell, L.C., and Sobell, M.B. Alcohol consumption measures. In: Allen, J.P., and Columbus, M., eds. *Assessing Alcohol Problems: A Guide for Clinicians and Researchers*. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism, 1995. pp. 55–73.

Sobell, L.C., and Sobell, M.B. Timeline Followback (TLFB) for Alcohol (Version 4.0b) [computer software]. Toronto: Addiction Research Foundation, 1996*a*.

Sobell, L.C., and Sobell, M.B. Timeline Followback Instructional Training Video for Alcohol [videotape]. Toronto, Canada: Addiction Research Foundation, 1996*b*.

Sobell, L.C., and Sobell, M.B. Alcohol Timeline Followback (TLFB). In: *Handbook of Psychiatric Measures*. Washington, DC: American Psychiatric Association, 2000. pp. 477–479.

Sobell, L.C.; Sobell, M.B.; Leo, G.I.; and Cancilla, A. Reliability of a timeline method: Assessing normal drinkers' reports of recent drinking and a comparative evaluation across several populations. *Br J Addict* 83:393–402, 1988.

Sobell, L.C.; Toneatto, T.; Sobell, M.B.; Leo, G.I.; and Johnson, L. Alcohol abusers' perceptions of the accuracy of their self-reports of drinking: Implications for treatment. *Addict Behav* 17:507–511, 1992.

Sobell, L.C.; Sobell, M.B.; Toneatto, T.; and Leo, G.I. What triggers the resolution of alcohol problems without treatment? *Alcohol Clin Exp Res* 17:217–224, 1993.

Sobell, L.C.; Toneatto, T.; and Sobell, M.B. Behavioral assessment and treatment planning for alcohol, tobacco, and other drug problems: Current status with an emphasis on clinical applications. *Behav Ther* 25:533–580, 1994.

Sobell, L.C.; Cunningham, J.C.; Sobell, M.B.; Agrawal, S.; Gavin, D.R.; Leo, G.I.; and Singh, K.N. Fostering self-change among problem drinkers: A proactive community intervention. *Addict Behav* 21(6):817–833, 1996.

Sobell, L.C.; Sobell, M.B.; Leo, G.I.; Agrawal, S.; Johnson-Young, L.; and Cunningham, J.A. Promoting self-change with alcohol abusers: A community-level mail intervention based on natural recovery studies. *Alcohol Clin Exp Res* 26:936–948, 2002.

Sobell, L.C.; Agrawal, S.; Sobell, M.B.; Leo, G.I.; Johnson-Young, L.; Cunningham, J.A.; and Simco, E.R. Comparison of the Quick Drinking Screen and the Timeline Followback with alcohol abusers. *J Stud Alcohol*, in press.

Sobell, M.B.; Sobell, L.C.; and VanderSpek, R. Relationships between clinical judgment, self-report and breath analysis measures of intoxication in alcoholics. *J Consult Clin Psychol* 47:204–206, 1979.

Sobell, M.B.; Sobell, L.C.; Klajner, F.; Pavan, D.; and Basian, E. The reliability of a timeline method for assessing normal drinker college students' recent drinking history: Utility for alcohol research. *Addict Behav* 11:149–162, 1986.

Sobell, M.B.; Bogardis, J.; Schuller, R.; Leo, G.I.; and Sobell, L.C. Is self-monitoring of alcohol consumption reactive? *Behav Assess* 11:447–458, 1989.

Straus, R., and Bacon, S.D. *Drinking in College*. New Haven, CT: Yale University Press, 1953.

Toneatto, T.; Sobell, L.C.; and Sobell, M.B. Predictors of alcohol abusers' inconsistent self-reports of their drinking and life events. *Alcohol Clin Exp Res* 16:542–546, 1992.

Tonigan, J.S.; Miller, W.R.; and Brown, J.M. The reliability of Form 90: An instrument for assessing alcohol treatment outcome. *J Stud Alcohol* 58(4):358–364, 1997.

Treatment Improvement Protocol Series 35 Consensus Panel. *Enhancing Motivation for Change in Substance Abuse Treatment*. Treatment Improvement Protocol Series No. 35. Rockville, MD: Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment, 1999.

Vuchinich, R.E.; Tucker, J.A.; Harllee, L.; Hoffman, S.; and Schwartz, J. Reliability of reports of temporal patterning. Paper presented at a poster session at the annual meeting of the Association for the Advancement of Behavior Therapy, Houston, TX, November 1985.

Vuchinich, R.E.; Tucker, J.A.; and Harllee, L.M. Behavioral assessment. In: Donovan, D., and Marlatt, G.A., eds. *Assessment of Addictive Behaviors: Behavioral, Cognitive, and Physiological Procedures*. New York: Guilford Press, 1988. pp. 51–93.

Webb, G.R.; Redman, S.; Sanson-Fisher, R.W.; and Gibberd, R.W. Comparison of a quantity-frequency method and a diary method of measuring alcohol consumption. *J Stud Alcohol* 51:271–277, 1990.

Webb, G.R.; Redman, S.; Gibberd, R.W.; and Sanson-Fisher, R.W. The reliability and stability of a quantity-frequency method and a diary method of measuring alcohol consumption. *Drug Alcohol Depend* 27:223–231, 1991.

Werch, C. Quantity-frequency and diary measures of alcohol consumption for elderly drinkers. *Int J Addict* 24:859–865, 1989.

Williams, G.D.; Proudfit, A.H.; Quinn, E.A.; and Campbell, K.E. Variations in quantity-frequency measures of alcohol consumption from a general population survey. *Addiction* 89:413–420, 1994.

Wyllie, A.; Zhang, J.F.; and Casswell, S. Comparison of six alcohol consumption measures from survey data. *Addiction* 89:425–430, 1994.