ELSEVIER

Journal of Substance Abuse 12 (2000) 1–21

Journal of Substance Abuse

Towards agreement on ways to measure and report drinking patterns and alcohol-related problems in adult general population surveys: the Skarpö Conference overview

Deborah A. Dawson[a,*], Robin Room[b]

[a]*National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Suite 514, Willco Building, 6000 Executive Boulevard, MSC 7003, Bethesda, MD 20892-7003, USA*
[b]*Centre for Social Research on Alcohol and Drugs, Stockholm University, Stockholm, Sweden*

Abstract

A thematic conference of the Kettil Bruun Society (KBS) for Social and Epidemiological Research on Alcohol was held in Skarpö, near Stockholm, on April 3–7, 2000. The goals of the meeting were to develop consensus sets of questionnaire items for measuring alcohol consumption and social harm, to delineate statistical and practical concerns related to the aggregation of consumption and harm data and to identify summary measures to be used for descriptive purposes and in analyses of the association between alcohol intake and alcohol-related outcomes. The results of the conference discussions are summarized below, with emphasis on both areas where the conference yielded recommendations for measures and methods of aggregation for analysis, and on areas where consensus could not be obtained and/or where additional research is needed.

Keywords: Consumption; Social harm; Measurement

1. Introduction

A thematic conference of the Kettil Bruun Society (KBS) for Social and Epidemiological Research on Alcohol was held in Skarpö, near Stockholm, on April 3–7, 2000. The Centre for Social Research on Alcohol and Drugs at Stockholm University hosted the meeting. Over 40 researchers from 12 countries with substantial social and epidemiological alcohol research

* Corresponding author. NIAAA/DBE, Suite 514, Willco Building, 6000 Executive Boulevard, MSC 7003, Bethesda, MD 20892-7003, USA. Tel.: +1-301-435-2255; fax: +1-301-443-8614.
E-mail address: ddawson@willco.niaaa.nih.gov (D.A. Dawson)

0899-3289/00/$ – see front matter
PII: S0899-3289(00)00037-7

traditions met to pursue an ambitious agenda designed to extend the state of the art in alcohol research. Specific goals included:

1. 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. reaching consensus as to the definition of social harm and developing a set of core items for measuring this overall construct as well as specific domains of harm; and
3. delineating statistical and practical concerns related to the aggregation of consumption and harm data, and identifying summary measures to be used for descriptive purposes and in analyses of the association between alcohol intake and alcohol-related outcomes.

In addition to the discussions of the 26 papers prepared for the meeting, many of which have been included in this issue, the conference took account of a draft World Health Organization document, International Guidelines for Monitoring Alcohol Consumption and Harm (World Health Organization, 2000), and of four draft instruments currently in preparation for international use.

This meeting built on the work of previous international conferences on related topics, stretching back to a 1986 conference of the International Group on Comparative Alcohol Studies in Poland on drinking patterns and drinking problems. This topic was re-addressed in two KBS-sponsored thematic meetings held in Toronto and Perth. Additionally, two workshops devoted to measuring and modeling alcohol consumption have been summarized in published overviews (Midanik & Harford, 1994) and proceedings (Dawson, 1998).

Considerable time was devoted at the conference to discussing possible international agreement on methods of measurement and analysis in alcohol surveys. While measurement issues for other types of epidemiological and surveillance data and in surveys of youth and of clinical populations were briefly discussed, the main emphasis in the discussions was on surveys of general adult populations. The results of the conference discussions are summarized below, with emphasis on both areas where the conference yielded recommendations for measures and methods of aggregation for analysis, and on areas where consensus could not be obtained and/or where additional research is needed. Specific recommendations are summarized in italics within each section.

2. Asking about drinking patterns

Conference participants agreed on the importance of developing and utilizing methods of asking about drinking that would more fully characterize both volume of intake and drinking patterns, especially those related to heavy drinking occasions. The discussion focussed on methods in studies where the emphasis is on characterizing the individual's drinking pattern and amount and relating these to other variables, e.g., as correlates or predictors of alcohol-related harm. It is specifically in this context where underestimation

of consumption may result in the attenuation of risk curves and misestimation of alcohol aetiological fractions, resulting in data that may not properly inform recommendations for low-risk drinking levels. Measures of current drinking patterns commanded most of the attention. This conference did not explicitly focus on lifetime measures of drinking pattern. Although most of the past-year measures arguably could be applied to period of heaviest drinking or specified decades of the lifetime, increased concerns about recall error have led to the use of methods incorporating more cognitive cues for estimating past drinking (see review in Lemmens, 1998).

2.1. Reference period for reporting

Those present agreed that different studies might have different purposes in asking about pattern and amount of drinking, and that the questions asked should be attuned to the purposes of the study. When a primary research goal is to link alcohol consumption with outcome measures, it is advantageous to use the same reference period for both to minimize problems of distinguishing cause from effect. Given the low prevalence of most alcohol-related problems, a past-year reference period is often optimal for this purpose. Alternatively, when the emphasis is on characterizing drinking occasions, investigators should consider using the ''Finnish approach'' — overall frequency of drinking followed by detailed questions about a period containing, at a minimum, the last four drinking occasions (e.g., past week for individuals drinking 4+ times a week, past month for those drinking once a week, etc.). Finally, when the sole interest is in characterizing drinking in the population as a whole, investigators might consider asking about drinking occasions on the two previous days, i.e., yesterday and the day before (Alanko, 1984). However, this approach of asking about the last one or two occasions is not a good method of establishing the amount and pattern of drinking at the level of the individual respondent (Skog, 1981). It also misses many infrequent drinkers, disproportionately comprised of women, youth, and the elderly, and it requires careful adjustment for possible seasonal effects (see also the recommendation by Medina Mora, Borges, & Villatoro, 2000, for repeated samples drawn at different times of the year to avoid this problem).

Recommendations:

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

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

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

Conferees strongly agreed that frequency of drinking should not be asked in an open-ended question requesting the exact number of times, e.g., number of times in the past year. In some

cultures, it is clear that respondents are less likely to answer "50 times" than to answer "once a week" (Ivis, Bondy, & Adlaf, 1997). Preferred options are:

1. to ask in terms of specified frequency ranges (e.g., twice a day, daily, 5–6 times a week/nearly every day, 3–4 times a week, 1–2 times a week, 2–3 times a month, once a month, 6–11 times a year, 1–5 times a year); or
2. to ask initially in terms of times per week (or per day in cultures where multiple drinking occasions per day are common), falling back on times per month if the respondent initially specifies "less than once a week" and times per year if "less than once a month."

Arraying the categories in order of descending frequency and starting with a frequency higher than most respondents will report should help to encourage respondents to think that their responses lie within accepted norms.

Recommendations:

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

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

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?

Those present generally agreed that the quantity of drinks consumed per occasion more clearly corresponds to the phenomenological reality of drinking than does the quantity of drinks consumed per day. Asking about quantity on a per day basis may also give a misleading impression of the blood alcohol level (BAL) associated with the quantity that is reported. However, there are likely to be big differences within and across cultures in what constitutes a drinking occasion for the respondent. Also, different research aims may dictate whether investigators want drinking occasions to be distinguished by the amount of time elapsed between drinks (if the emphasis is on the implications related to BAL), or by shifts in drinking partners and venue (if the emphasis is on the influence of drinking context). Thus, the conference recommendation is that quantity be asked in terms of number of drinks per day, with the justification being that it is a more "objective" measure. However, some of those present cautioned that a drinking day may span more than one calendar day (i.e., when drinking continues after midnight) and that questions need to be worded to take this into account and not have quantities consumed on occasions such as these divided between 2 days or reported twice.

Recommendations:

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"?

In graduated frequencies (Greenfield, 2000) and other approaches that ask about the frequency of consuming specified quantities (e.g., 5+ drinks, largest number of drinks, etc.), the series of questions can start with lower quantities and work up or start with higher quantities and work down. One well-established approach (Greenfield, 2000) asks first about the largest quantity consumed in the past 12 months, and uses that to set the highest quantity and then work down. Many of those present felt that a "working-up" approach might result in more underreporting but acknowledged that there are at present too few methodological studies to provide the basis for recommending one of these approaches over the other.

Recommendations:

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

The following quantity thresholds, or some approximation thereof that includes at minimum the 60-g cutoff, were recommended for graduated frequency questions and for questions about specific amounts consumed in the past 12 months or in the past month:

240 g ethanol (usually only for largest-amount questions)
144 g ethanol
96 g
60 g
(36 or 40 g)
(20 or 24 g)

One or both of the lower parenthetical thresholds is recommended if volume of drinking is to be estimated from these questions. If the main emphasis is on measurement of heavy drinking, these thresholds could be omitted (however, see Section 2.6 for a discussion of using these thresholds to provide for different definitions of heavy, hazardous, or harmful drinking for men and women).

Questions should not be asked in terms of grams of ethanol, but in terms of the relevant equivalents for the particular population being studied, usually standard drinks for which examples have been provided (e.g., "the equivalent of a 12-oz beer, a 4-oz glass of wine, or a drink containing 1 oz of distilled spirits"). The size of a standard drink typically used in surveys and analyses varies across cultures (International Centre for Alcohol Policies, 1998), e.g., 10 g in Australia and 12–14 g in the United States. The 8-g "standard unit" in British surveys is also often misinterpreted in terms of "drinks." The 60-g cutoff, which corresponds to the definition of hazardous consumption (>60 g) proposed in the International Guidelines for Monitoring Alcohol Consumption and Harm (World Health Organization, 2000), roughly corresponds to the heavy drinking measure of 5+ drinks used in US studies and 6+ drinks used both in Australian studies and the AUDIT, which also assumes a standard drink of 10 g.

The equivalence between ethanol and standard drinks, regardless of size, is problematic. If a person starts but does not complete a fifth drink, does this count towards a question that asks for the frequency of drinking five (or 5+ or 5–7, etc.) drinks? This concern is not as trivial as it may seem, because when people are asked to convert their actual intake (which may be consumed in decidedly nonstandard-sized units) into a number of standard drinks, the result often will not be a round number. The questionable ability of individuals to perform this conversion, because of either the mathematics involved or lack of knowledge of their own drink sizes (see Kaskutas & Graves, 2000, for evidence of how great the magnitude of these problems may be), further adds to the inexactness of these equivalencies. The conference did not reach any consensus on how to deal with these problems, beyond recommending that the interviewer should have an available table of local equivalencies (in bottles, liters, etc. of actual beverages) for each threshold used. Depending on the interview mode, a simplified version might also be made available to the respondent.

Participants noted that responses in this area are approximate and that analyses should not attribute a spurious exactness to them. There was agreement on the need for methodological studies on improving measurement on this dimension, which will need to be carried out specifically for each given cultural situation (and may need to be repeated periodically as circumstances change). In cultures where questions are usually asked in terms of "drinks," there is a need for research on what respondents mean by "a drink" of different beverages and in different circumstances (e.g., in a bar or at home).

Recommendations:

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?

There was disagreement among those at the conference about whether and under what circumstances different consumption thresholds should be set for men and women. This debate entails two important and quite separate issues: (1) whether men and women experience different effects, particularly acute effects, at the same level of consumption, and (2) the extent, if any, to which these differences should be reflected in the questions used to assess intake.

Regarding the first of these issues, the argument in favor of a lower threshold for acute effects among women is based on differences in average blood alcohol attained by women and men from a given number of drinks in a given period of time (Dawson & Archer, 1992), and in-the-event risk analyses that have compared men and women (McLeod, Stockwell, Stevens, & Phillips, 1999). The argument against a lower threshold is supported by other studies indicating that women do not seem to get in more trouble than men for a given amount of drinking, possibly because of differences in drinking rates, as well as behaviour after drinking (see review in Graham, Wilsnack, Dawson, & Vogeltanz, 1998).

These findings suggest that different thresholds for men and women may be more defensible if used to obtain rough equivalence of acute physical and cognitive effects than if used to obtain equivalence of acute behavioral and social effects. However, conference participants agreed that more research is needed to explain the contradictory research findings in this area. Important areas for consideration include possible gender differences in drinking rates and the frequency of drinking with meals, in the underlying probabilities of engaging in behaviors, such as fighting and driving, especially driving on the types of social occasions where alcohol is likely to be consumed, and in the quantities of drinks consumed on heavy drinking occasions.

Regarding the second issue, if a series of threshold-type questions (e.g., graduated frequencies) is used to estimate volume of intake, there are methodological advantages to keeping the thresholds (or the quantity categories implied by these thresholds) as comparable as possible for men and women. Even if the thresholds corresponding to hazardous drinking are defined so as to vary for men and women (e.g., 60 and 40 g, respectively), this can be accommodated within the framework of a single series of thresholds, simply by including all the male and female thresholds in that series. Moreover, some present argued that even if different "low-risk" drinking guidelines are desirable for men and women, either for acute or chronic consequences of drinking, that these should be supported by analyses showing different levels of risk associated with given levels of intake, derived from a single set of thresholds, rather than by building the assumption of gender differentials in risk into the questions used to assess this relationship.

Recommendations:

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?

Two methods have been used for asking about drinking above particular thresholds:

1. discrete quantity bands: How often have you had 8–11 drinks? How often 5–7 drinks? How often 3–4 drinks? etc.
2. cumulative quantities: How often have you had eight or more drinks? How often five or more drinks? How often three or more drinks? etc.

In the present state of knowledge, the conferees recommended using cumulative quantities in questionnaires intended for cross-cultural use or comparison. A cumulative quantity question requires spelling out just one quantity threshold in each question, which is likely to be less cumbersome and confusing to adapt cross-culturally.

With this approach and a "working down" strategy, it was recommended that the respondent be reminded that each response should include the drinking days, which had

already been reported with higher quantities. There is a need for methodological research on whether and how much these different strategies of questioning yield different results.

Recommendations:

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

Volume of drinking can be computed from just two questions, asked for all alcoholic beverages together: frequency of drinking, and usual quantity of drinking. However, it has long been recognized that this would yield a considerable underestimate if a respondent drank small amounts daily but large amounts on a weekend; logically, and assuming that usual quantity refers to the modal rather than average amount, his or her responses should not reveal anything about the weekend quantity (Room, 1990). Nevertheless, the conference recommended asking a single usual-quantity question, along with the other questions recommended, to offer comparability with the wide range of studies that have asked it.

Coverage of consumption with usual-quantity questions is somewhat improved if both the frequency and usual quantity are asked for specific beverage types — for instance, if the respondent drinks beer on weekdays but spirits on weekends. Such an approach gets useful data on beverage choice with relatively few questions. However, better results will usually be obtained in computing volume of drinking by using a graduated-frequencies approach (Rehm, Greenfield, et al., 1999), or at least by adjusting the volume estimate from "usual quantity" upwards on the basis of responses to one or more "graduated frequency" or frequency of heavy drinking questions (Dawson, 1998; Polich & Armor, 1982).

Recommendations:

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

There was agreement that it was desirable for most study purposes to collect some data about beverage choice, but this might be limited, e.g., to asking for the frequency of drinking of each beverage, or to asking which beverage is most frequently consumed. The customary division into wine, beer, and spirits is problematic for computing volume of drinking even in the European and North American cultures in which it has been customary (e.g., resulting in underestimation of the ethanol obtained from fortified wines and strong beer/malt liquor and overestimation of the ethanol obtained from light beer). In a wider range of societies, a wider range of types of beverage needs to be used. Even in North American studies, the assumption that coolers can be counted as wine has been invalidated by the widespread introduction of spirits and malt based coolers.

Recommendations:

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.

The types of beverages included must vary to reflect individual countries' consumption patterns.

2.10. More precise measurement of indicators of attained BALs

Where studies desire to improve the precision of measurement of attained BALs, further measurements will be needed. These include:

1. duration of drinking occasions. This information is easily collected concerning recent drinking occasions, as in the "Finnish method" of asking questions (Alanko, 1984). There is less experience with collecting it with questions asking about customary quantities. It was suggested that a follow-up question ("Over how many hours was that, usually?") could be asked for each discrete quantity level or usual-quantity question. Alternatively, duration questions could be asked for each context in a drinking context series (see below).
2. Data for estimating total body water or body mass index. Body weight needs to be asked to estimate BAL by methods such as have used in Finnish studies (Bruun, 1969) to estimate times the respondent has exceeded a given BAL. A more refined estimate, using body mass index, also requires asking height. The optimum measure of total body water (e.g., Watson, Watson, & Batt, 1980) requires only the addition of age and gender, items that typically will be collected as a matter of course. Even adjustment for total body water will yield only an approximation of the BAL associated with a given level of ethanol intake, in that numerous other factors, such as the duration of the drinking occasion and food intake, also affect the BAL.

Recommendations:

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 number of drinking context differentiations were identified as potentially important in cross-cultural studies:

with meals/without
on a weekday/on a weekend
in public (bars and restaurants)/not in public
alone/with others (possibly including detail on who those others were, e.g., family members, friends, work colleagues, etc.)

Those present noted that asking about contexts of recent drinking occasions (with whom; in what location; what was the nature of the occasion) was an efficient method of tracking drinking contexts and changes in contexts at a population level. A series of questions about different drinking contexts, used in North America, asks about the frequency of being in that context, the frequency of drinking in that context, and the usual quantity in that context. Such a series is recommended for detailed studies of drinking patterns that do not use a "recent-occasions" methodology. The series has been also used as an alternative method of computing volume of drinking (Single & Wortley, 1994). When used for this purpose, it has yielded coverage rates as high as 65%, although it is difficult to ensure that the contexts are mutually exclusive (Wylie, Zhang, & Caswell, 1994) or fully comprehensive.

Recommendations:

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

2.12. Frequency of getting drunk/intoxication

There is a long history of asking questions on the frequency of drunkenness or something comparable. Questions of this type often have been interpreted as alternative ways of asking about heavy drinking. Specific questions that have been used internationally include how often the respondent "drank enough to feel the effects" (for Ontario respondents asked how many drinks it took to feel the effects, "three or four drinks" was the modal response, Ferris, Templeton, & Wong, 1994), or drank enough to feel "drunk" or "intoxicated."

Conference participants felt that asking about frequency of drunkenness or intoxication was not a satisfactory method of measuring frequency of heavy drinking per se, since the concept of being "drunk" includes a large component of cultural definition (e.g., as per Midanik's 1999 study in which respondents reported that it took fewer drinks to feel drunk or feel the effects of alcohol during "drier" as compared to "wetter" time periods). However, they felt it was important to ask about the culturally influenced experience of being drunk as a separate variable from amount of drinking, and most thought it preferable to ask about being "drunk" rather than about "feeling the effects" of alcohol, as the latter may be less clearly influenced by culture. (Though not discussed at the meeting, the logical corollary to this is that researchers wanting to minimize gender, ethnic, or other cultural differences in the interpretation of drunkenness or intoxication within their survey population may wish to opt for a question on "feeling the effects" or, alternatively, to provide some specific examples of how drunkenness or intoxication are defined, e.g., "that is, times when your speech was slurred, you felt unsteady on your feet, or you had blurred vision.")

Recommendations:

Questions on frequency of drunkenness/intoxication are preferable to those on feeling the effects.

Although valuable 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

For studies whose intent is to characterize individuals' drinking levels and patterns, participants recommended asking about the following when only a very limited number of questions can be asked:

Recommended items:

abstention — lifetime and past 12 months
overall frequency of drinking (all alcoholic beverages together)
usual quantity of drinking (all alcoholic beverages together)
frequency of consuming >60 g ethanol in a day
(1st alternative: if usual quantity was >60 g, ask frequency of consuming >96 g; 2nd alternative: largest amount drunk in a day in the past 12 months and how often that amount was consumed)
frequency of drunkenness (if possible)

2.14. Recommended set of questions on drinking amount and pattern

For studies that are able to include a larger number of questions, some or all of the following items are recommended. Within the context of this expanded set of items, those that are desirable but of lower priority are preceded by an asterisk:

Recommended items:

abstention — lifetime and past 12 months
largest amount drunk in last 12 months (maximum quantity), all beverages together
graduated frequencies questions, all beverages together:
*cut-offs: * ≈ 24 and/or ≈ 36, 60, 96, 144, 240 g for largest amount (less desirable alternative: frequency of drinking >60 g)*
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)
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)
beverage-specific usual quantities of drinking
beverage-specific size of usual drink
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*

The items in this list lend themselves to two general approaches to measuring consumption. One is the estimation of volume and pattern from graduated frequencies questions for all beverages combined, asked in terms of standard drinks, with data on the frequency (and possibly usual quantities) of drinking specific beverages gathered both as an alternative estimate of (or check on) volume and to allocate total intake among beverages. The second is the estimation of volume by means of detailed beverage-specific questions that include self-reported drink size and beverage-specific maxima and their frequencies, with pattern of intake determined from questions on overall frequency of drinking and frequencies of drinking >60 g (and possibly maximum quantity) in terms of standard drinks for all beverages combined. These two approaches need not be mutually exclusive, but limits in funds, time, and respondent willingness make it unlikely that both could be fully utilized in a single survey instrument.

Although conference participants offered ad hoc comments on advantages and limitations of each of these approaches, the two were not formally compared and there was no consensus as to which was superior. In opting for one approach over the other, factors to be considered include the number of types of beverages for which beverage-specific questions would be required, the ability of respondents to convert their drinks into standard drinks, and whether the improvement in volume estimation that might come from allowing respondents to report their actual drink sizes is outweighed by the advantages of having pattern data that are standardized into prespecified rather than respondent-specified quantity levels. With respect to the latter, the mode of interview is a consideration, since respondent aids to estimating usual drink sizes (actual vessels or life-sized photographs) cannot be used in telephone interviews.

3. Aggregating drinking patterns for analysis

There is a need for conceptual clarity and to avoid colinearity problems in multivariate analysis. Conference participants identified the following possible alternative sets of drinking-pattern dimensions to be used in multivariate analyses (but see also Gmel, Rehm, Room, & Greenfield, 2000 and Rehm & Gmel, 2000 for a more detailed discussion of the concerns that guide the selection of analytic exposure measures, and Dawson, 2000 for an empirical comparison of some alternative combinations):

Recommended sets of items:

(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; Greenfield, 1986)

(C) Frequency of drinking
Usual/Average quantity per occasion
Variance of quantity or frequency of 5+, etc.

Although there was little discussion of the relative merits of these combinations, there are many considerations that might lead to the choice of one over another. Combination A offers the advantage of having a measure of variability that is constant and can be interpreted the same way for all respondents, reflecting the acute blood alcohol and/or behavioral effects associated with a given level of per-occasion intake. Combination B offers a measure of variability that is a function of the overall or mean level of intake, thus reflecting the acute effects of deviation from the level of intake for which some degree of tolerance may have been developed. Combination C allows the effects of quantity and frequency (overall, not just of heavy drinking) to be distinguished, but the addition of a third variable reduces the efficiency of the model and makes the estimation of interactions with other variables, e.g., sociodemographic variables, more cumbersome.

This last issue was also identified by participants as an argument against using multiple categorical consumption variables, such as those that have been derived from the Volume–Variability Index (Cahalan, 1969; Cahalan, Cisin, & Crossley, 1968; see also extensions of this typology in Caetano & Herd, 1984; Cahalan, Roizen, & Room, 1976; Wallack, 1978). On the other hand, such typologies offer the advantage of being very easy to interpret, especially when compared to the mathematical transformations that frequently are required to make continuous measures, such as those in A, B, and C, satisfy the linearity assumption of multiple regression models.

Recommendations:
For summary measures of alcohol consumption, analytic goals should guide the choice between multiple categorical variables and combinations of unidimensional variables.

4. Measuring social problems in relation to drinking

The discussion of alcohol-related consequences was limited primarily to the arena of social problems from drinking. Although conference participants included some aspects of health under the rubric of social harm, the discussion did not cover dependence and other psychiatric conditions or mental states related to drinking. Those present noted that the DSM-IV "alcohol abuse" category (American Psychiatric Association, 1994) includes items in the area of social harm but felt that this category is not conceptualized or operationalized in broad enough terms to serve as a measure of social problems from drinking. Despite the important policy implications of any social and/or health benefits of drinking, these were outside the range of issues considered at this meeting. Rather, the focus was on those consequences of drinking

that have been among the most difficult to measure adequately but that also are of great importance to policy makers, e.g., in terms of reducing drinking–driving incidents, accidents and injuries, domestic and other violence, family instability and hardships, and damage to productivity and employment.

Researchers attending the conference agreed that there is a need to develop and test measures of social harm from drinking, both for harm resulting from the respondent's own drinking, and for harm resulting from others' drinking. Such measures are needed at the level of particular problem areas, and also at more general levels of social harm or tangible consequences. The issue of whether health problems should be aggregated with social problems or kept separate was discussed but not resolved.

Testing is required to establish both the reliability and validity of social harm measures. Reliability should be investigated by means of (1) test–retest studies separated by an interval of 2–4 weeks, (2) split half-sample comparisons if sample size permits, and (3) calculation of inter-item consistency. All of these techniques are ways of establishing that the measures of social harm are measuring the same thing over time, among different people, and across different items. With respect to the latter, a low level of consistency among items need not necessarily indicate lack of reliability or validity for the individual items. If adverse social consequences are multidimensional in nature, individual items or groups of items may have reliable and valid self-reporting without resulting in high internal consistency for all items taken as a whole; however, this might argue against combining all the items into a single, unidimensional summary scale or score.

The construct, face, and predictive validity of social harm measures, i.e., the extent to which they mean what we want them to mean, should be established by means of (1) verification of self-reports against external sources, e.g., registry data, police statistics, etc., (2) testing correlation among groups of items that theoretically ought to reflect the same underlying dimension of social harm, and (3) quasi-longitudinal examination of likely outcomes, e.g., whether individuals reporting alcohol-related problems with partners are more likely to become divorced.

Most social problems from drinking are inherently interactional, involving not only drinking and associated behaviour defined as problematic but also a reaction or response from another or from others (see also Room, 2000). There is a need for study design that takes account of the interactional nature of the problems, e.g., questioning not only the drinker but also his/her family members or social network.

There also is a need to study processes of attribution of problems to drinking, both for problems related to own drinking and for problems related to others' drinking. The studies need to recognize variations in the type and extent of attribution. The relation between the drinker's attribution and others' attributions, and between global attributions and attributions to events, should be studied from the perspective of item development and aggregation, as well as substantively. In a cross-cultural frame, the influence of cultural expectations on attributions and on problem reporting needs to be examined.

Measures of interpersonal problems from drinking should include measurement both of problematic behaviour and of responses, with attention to intensity of interaction and reaction, and alcohol and nonalcohol attributions of problems. In general, there is a need for studies that go deeply into particular problem areas (e.g., family, or work, or

social relationships) and that look at the interrelationship of multiple problem indicators in that domain.

Recommendations:

Measures of social harm should examine both harm resulting from the respondent's own drinking and from others' drinking.

Measures of social harm need to be tested for reliability and validity and in terms of the interrelationship of multiple indicators within individual problem domains.

The process of attribution of problems to drinking needs further study, including the role of cultural expectations on the attribution of harm to alcohol.

4.1. Problems related to own drinking

Conference participants noted that questions about problems related to one's own drinking tended to fall into a few types:

Global self-judgement — such as the questions about harm to different life-areas (Rehm, Frick, & Bondy, 1999). These items have a relatively high intercorrelation, reflecting that they all presuppose an acknowledgement of one's drinking as causing problems. "Health effects" tend to be the most frequently reported, but this is the most problematic item, since it is often fears or transient (hangover) symptoms that are being reported (Greenfield & Rogers, 1997).

Others' judgement (as reported by the respondent) — these are in principle externally verifiable (the others could be asked themselves). This is the area in which, traditionally, severity differentiations have been made in asking questions (spouse left or kicked you out because of your drinking; spouse angry because of your drinking; spouse talked to you about your drinking). There is an attribution to drinking being made by the other, and also to a varying extent by the respondent in reporting it. It was noted that the most problematic drinkers may eventually receive less social control from others — the others have given up on expecting and trying to promote any change, and may in fact have cut-off relations.

"Objective" indicator — such as an arrest for drinking–driving, going to work with a hangover, etc. As indicators of social problems, many of these items have the difficulty that the indicator is not necessarily in itself a social problem.

Recommendation:

Studies should be undertaken to examine the interrelations of the above types of items, in the context of the different major social roles (family, work, citizen) in which they play out. This is both a methodological and a substantive agenda.

4.2. Attribution to alcohol

Conference participants noted that measures of social harm could reflect at least three levels of alcohol attribution: (1) items directly attributed to alcohol, e.g., "lost a job because of my drinking"; (2) items temporally linked to drinking but with no direct causal attribution,

e.g., "got into a fight after drinking"; and (3) items not attributed or linked to drinking in any way, e.g., "had an injury that required medical attention". Most alcohol surveys traditionally have included items of the first two types, but few have included items of the third type. An area of strong consensus was that studies should consider asking parallel questions about the experience of problems with and without mention of or attribution to drinking, preferably asking the nonattributed items before the comparable alcohol-attributed items. The Addiction Severity Index (ASI), widely used in clinical studies of alcohol problems (McLellan et al., 1992), was offered as a possible starting point for such unattributed problem questions; however, many of its scales lie beyond the bounds of social harm and both its mode of administration, its severity ratings (based on the interviewer's rating of need for additional treatment) and possibly the time frame for the questions would require modification for use in general population surveys.

One of the advantages of including unattributed items would be that these items could be asked of nondrinkers, as well as drinkers, providing a reference group comparable to what is typically used in epidemiological studies of chronic health outcomes. Another advantage is that items that do not ask respondents to determine whether alcohol was the cause of a problem are likely to be more reliable and more comparable across cultures than those involving attribution. On the other hand, when items are not directly attributed to alcohol use, their temporal order and contiguity with respect to drinking may be impossible to determine with certainty.

Recommendations:

Studies should ask parallel questions about the experience of problems with and without attribution to drinking, preferably asking the nonattributed items before the comparable alcohol-attributed items and asking the nonattributed items of nondrinkers, as well as drinkers.

4.3. Estimating the cost of social harm

There is a need to develop measures of social harm from drinking that can be used in economic-costing studies. Consideration should be given in aggregating measures to possible boundaries of "externalities" in economic analyses. Willingness-to-pay approaches should be explored. Participants noted that cost estimates, when they can be made, do offer a continuous measure of severity of the problem. The ranking of costs may be meaningful even if the absolute costs reported are not. Participants suggested the use of rank ordering of problems as a way of at least arriving at an ordinal measurement. "How much did you care?" and "How much did others care?" are possible ways of arriving at rankings of importance.

Those present discussed the relationship between problem self-attribution questions and economically oriented studies (e.g., of cost of illness) from several angles and suggested that problems need some external measurement — not just self-perceived harm — to justify public expenditures for alcohol problems. Although they noted that there are harms without money value that need to measured, they recommended that at the least surveys could ask about items for which costs are available: days lost from work, emergency room visits, etc.

Recommendations:

If possible, surveys should include social harm items for which costs are available.

Studies are needed to examine approaches, such as willingness to pay, for ranking the "costs" or severity of social harms for which costs are not available.

4.4. Problems related to others' drinking

Participants recommended that consideration also be given to developing and validating parallel sets of items about problems due to one's own drinking and problems due to others' drinking. In the area of measurement of victimization by others' drinking, they suggested that the alcohol literature should follow the lead of the victimization literature in classifying and asking about problems. In this area, there was some feeling that items should in principle be verifiable through external social statistics. It was noted that violence in the home and outside the home should be asked about separately.

Recommendations:

Studies should consider including parallel questions on harm resulting from own and others' drinking, distinguishing violence inside and outside the home.

4.5. Areas to be covered in measuring problems related to drinking

There was a brief discussion of possible criteria for choosing items for inclusion. One suggestion was that the emphasis should be on the more prevalent harms and on those that distinguish men and women. The following areas were enumerated:

marital and partner problems
problems with family, children, parents
problems with friendships and social life
work (school) problems
financial problems (including spending money needed for essentials)
health problems
casualty and injury
criminal behaviour, police responses
drinking–driving and other criminal behaviour risking casualties
fighting and violence
sexual misbehaviour
risk-taking behaviours (e.g., risky sexuality, riding with a drunk driver)
spiritual well-being

4.6. Minimum set of questions on problems due to drinking

The following are suggested as a minimum set of items on social problems related to drinking for use internationally in surveys. Items preceded by an asterisk are of secondary priority.

Recommended list of problems from own drinking:

involved in accident when self had been drinking
involved in fight when self had been drinking
harm to family or home life from own drinking
harm to employment or work life (or schoolwork) from own drinking*
harm to friendships or social life from own drinking
** harm to finances from own drinking*
** harm to health from own drinking*
** harm to outlook on life/spiritual well-being from own drinking*
** behaved while drinking in a way you regretted afterward*

Recommended list of problems from others' drinking (Consider asking for some or all items: Do you know the person? Who is it?)

involved in accident due to someone else's drinking
been pushed, hit, or assaulted by someone who had been drinking
family problems or marriage difficulties due to someone else's drinking
had friendships break up as a result of someone else's drinking
have been harassed or bothered by someone who had been drinking
** your work disrupted or your job prospects hurt by someone else's drinking*

Recommended list of nonattributed problems

** Addiction Severity Index items for law, family, work, social life areas*
** how satisfied with health, with quality of life*
** involvement in accident, involvement in fight*

5. Aggregation of social harm measures for analysis

By virtue of falling at the end of the conference, the important question of how to combine measures of social harm into summary measures, such as scales or scores, was barely addressed, except in the paper by Gmel et al. (2000). Conference participants did agree that additional research was needed in the following areas before specific summary measures could be proposed.

Recommendations:

Studies must be conducted to determine whether social harm is inherently unidimensional or multidimensional.

Studies are needed to determine whether this dimension or these dimensions can be represented by simple counts of problem items endorsed, implying that all harms within a given dimension should receive an equal weight.

Further research is needed to determine whether severity can be fully accounted for by such a count (weighted or unweighted) or only by additionally considering the frequency with

which different types of harm occur or their inherent severity, e.g., being left by your partner vs. your partner complaining about your drinking.

6. Concluding remarks

This conference represented a major step forward in developing international consensus on the important issues of how to measure, aggregate, and link alcohol consumption and alcohol-related harm. The papers that follow are among those that helped to stimulate the discussion of these issues. Broadly speaking, these papers fall into several categories: (1) overviews of the types of questions that have been used to measure alcohol consumption and social harm; (2) reviews and empirical analyses of specific topics that include graduated frequencies, daily reporting of consumption, estimation of standard drinks, and hazardous consumption; (3) reviews of the issues surrounding the aggregation of consumption and harm measures and the resulting impact on multivariate models linking intake and problems; and (4) reviews of three individual countries' measurement traditions, which illustrate how much these measures vary, often in order to reflect the realities of different types of drinking cultures. These papers provide an excellent point of reference for the additional research.1 that is needed to resolve those questions that were not fully addressed or answered during the course of this conference.

Acknowledgments

The authors express their gratitude and appreciation for the many valuable contributions of Richard Wilsnack, Tim Stockwell, Tom Greenfield, and Florence Kellner in the preparation of this conference overview.

References

Alanko, T. (1984). An overview of techniques and problems in the measurement of alcohol consumption. *Research Advances in Alcohol and Drug Problems*, *8*, 209–226.

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

Bruun, K. (1969). The actual and the registered frequency of drunkenness in Helsinki. *British Journal of Addiction*, *64*, 3–8.

Caetano, R., & Herd, D. (1984). Black drinking practices in northern California. *Journal of Alcohol and Drug Abuse*, *10*, 571–587.

Cahalan, D., Cisin, I.H., & Crossley, H.M. (1969). A multivariate analysis of the correlates of drinking-related problems in a community study. *Social Issues*, *17*, 234–247.

Cahalan, D., Cisin, I. H., & Crossley, H. M. (1968). American drinking practices; summary of findings from a national probability sample: II. Measurement of massed versus spaced drinking. *Quarterly Journal of Studies on Alcohol*, *29*, 642–656.

Cahalan, D., Roizen, R., & Room, R. (1976). Alcohol problems and their prevention: public attitudes in California. In: R. Room, & S. Sheffield (Eds.), *The prevention of alcohol problems* (pp. 354–403). Sacramento: State Office on Alcoholism.

Dawson, D. (Ed.) (1998). Proceedings, international workshop on consumption measures and models for use in policy development and evaluation. *Alcoholism: Clinical and Experimental Research*, *22* (2), 1S–81S (Supplement).

Dawson, D. (2000). Alternative measures and models of hazardous consumption. *Journal of Substance Abuse* (this issue).

Dawson, D., & Archer, L. (1992). Gender differences in alcohol consumption: effects of measurement. *British Journal of Addiction*, *87*, 119–123.

Ferris, J., Templeton, L., & Wong, S. (1994). *Alcohol, tobacco and marijuana: use norms, problems and policy attitudes among Ontario adults*. Toronto: Addiction Research Foundation (Research Document 118).

Graham, K., Wilsnack, R., Dawson, D., & Vogeltanz, N. (1998). Should alcohol consumption be adjusted for gender differences. *Addiction*, *93*, 1137–1147.

Gmel, G., Rehm, J., Room, R., & Greenfield, T. (2000). Dimensions of alcohol-related social harm in survey research. *Journal of Substance Abuse* (this issue).

Greenfield, T. K. (1986). Quantity per occasion and consequences of drinking: a reconsideration and recommendation. *International Journal of Addiction*, *21*, 1059–1079.

Greenfield, T. K. (2000). Ways of measuring drinking patterns, and the difference they make: experience with graduated frequencies. *Journal of Substance Abuse* (this issue).

Greenfield, T. K., Rogers, J. D. (1997). What's in a problem? Type and seriousness of harmful effects of drinking on health. Berkeley, Alcohol Research Group. Working paper F405.

International Center for Alcohol Policies. (1998). What is a "standard drink"? Washington, ICAP, Report No. 5 (on web at: http://www.icap.org/icapreport5.html).

Ivis, F. J., Bondy, S. J., & Adlaf, E. M. (1997). Effect of questions structure on self-reports of heavy drinking: closed-ended versus open-ended questions. *Journal of Studies on Alcohol*, *58*, 622–624.

Kaskutas, L., & Graves, K. (2000). Alternative to standard drinks as a measure of alcohol consumption. *Journal of Substance Abuse* (this issue).

Lemmens, P. (1998). Measures of lifetime drinking history. *Alcoholism: Clinical and Experimental Research*, *22,* (2), 29S–37S (Supplement).

McLellan, A. T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., & Argeriou, M. (1992). The fifth edition of the addiction severity index. *Journal of Substance Abuse and Treatment*, *9*, 199–213. Available at: http://silk.nih.gov/silk/niaaa1/publication/instable.htm.

McLeod, R., Stockwell, T., Stevens, M., & Phillips, M. (1999). The relationship between alcohol consumption patterns and injuries. *Addiction*, *94*, 1719–1734.

Medina Mora, M. E., Borges, G. G., & Villatoro, J. (2000). The measurement of alcohol consumption and harm in Mexico. *Journal of Substance Abuse* (this issue).

Midanik, L. M. (1999). Drunkenness, feeling the effects and 5 plus measures. *Addiction*, *94*, 887–897.

Midanik, L. T., & Harford, T. C. (1994). Alcohol consumption measurement: introduction to the workshop. *Addiction*, *89*, 393–394.

Polich, J. M., & Armor, D. A. (1982). Measurement of alcohol consumption. In: E. Pattison (Ed.), *Encyclopaedic handbook of alcoholism* (pp. 72–80). New York: Gardiner Press.

Rehm, J., Frick, U., & Bondy, S. (1999). Reliability and validity analysis of an alcohol-related harm scale for surveys. *Journal of Studies on Alcohol*, *60*, 203–208.

Rehm, J., & Gmel, G. (2000). Aggregating dimensions of alcohol consumption to predict consequences. *Journal of Substance Abuse* (this issue).

Rehm, J., Greenfield, T. K., Walsh, G., Xie, X., Robson, L., & Single, E. (1999). Assessment methods for alcohol consumption, prevalence of high risk drinking and harm: a sensitivity analysis. *International Journal of Epidemiology*, *28*, 219–224.

Room, R. (1990). Measuring alcohol consumption in the United States: methods and rationales. *Research Advances in Alcohol and Drug Problems*, *10*, 39–80.

Room, R. (2000). Concepts and items in measuring social harm from drinking. *Journal of Substance Abuse* (this issue).

Single, E., & Wortley, S. (1994). Comparison of alternative measures of alcohol consumption in the Canadian National Survey of alcohol and drug use. *Addiction*, *89*, 395–399.

Skog, O.-J. (1981). Distribution of self-reported alcohol consumption: comments on Gregson and Stacey. *Psychological Reports*, *49*, 771–777.

Wallack, L. (1978). *An assessment of drinking patterns, problems, knowledge and attitudes in three Northern California communities*. Berkeley: Alcohol Research Group (Report C21).

Watson, P. E., Watson, I. D., & Batt, R. D. (1980). Total body water volumes for adult males and females estimated from simple anthropometric measurements. *American Journal of Clinical Nutrition*, *33*, 27–39.

World Health Organization. (2000). *International guide for monitoring alcohol consumption and related harm*. Geneva: Department of Mental Health and Substance Dependence, Noncommunicable Diseases and Health. WHO Publication Number WHO/MSD/MSB/00.4.

Wylie, A., Zhang, J.-F., & Casswell, S. (1994). Comparison of six alcohol consumption measures from survey data. *Addiction*, *89*, 425–430.