



Alcohol Intake Assessment: The Sober Facts

Gerda I. J. Feunekes, Pieter van 't Veer, Wija A. van Staveren, and Frans J. Kok

Recent recommendations in regard to the level of alcohol intake have mainly been based on epidemiologic studies which relied on self-reported amounts of alcohol consumed. Therefore, it is important to be aware of the quality of self-reported measures of alcohol intake. Alcohol intake assessment methods were reviewed with respect to their capacity to rank individuals according to alcohol intake and their ability to explain the variation in the level of intake in population samples. In 33 methodological papers published after 1984, alcohol intake was assessed by five main methods: quantity frequency, extended quantity frequency, retrospective diary, prospective diary, and 24-hour recalls. The mean level of alcohol intake differed by 20% between these methods. It was also found that when researchers asked specifically about intake of beer, wine, and liquor, this resulted in 20% higher estimates of intake. These percentages were similar among populations with low and high mean alcohol consumption (4 vs. 10 drinks per week). It was found that ranking of individuals according to intake was satisfactory, with weighted correlation coefficients between methods ranging from 0.63 to 0.73. The authors conclude that, when there is sufficient evidence that alcohol intake is underestimated in a population, methods that enquire about both the frequency and amount consumed, for beer, wine, and liquor, separately, will yield the most realistic levels of intake. *Am J Epidemiol* 1999;150:105–12.

alcohol drinking; alcoholic beverages; alcoholic intoxication; epidemiologic methods

Recent liberalization of recommendations on alcohol intake, e.g., in Great Britain, toward 2–3 drinks per day for women and 3–4 drinks per day for men (1), has been based mainly on epidemiologic studies which have relied on self-reports of “usual” alcohol intake of individuals. This underlines the importance of the quality of self-reported alcohol intake measures. These epidemiologic studies primarily aimed to assess the strength of the association between alcohol and health, with a focus on ranking individuals according to intake, rather than on accurately assessing the absolute level of intake. Different assessment methods, however, may give systematic differences in mean level of alcohol intake. To advise the public on “sensible” limits of alcohol intake and to be able to verify compliance, methods are needed that properly *rank* individuals according to alcohol intake, and that also assess correctly the absolute *level* of intake.

Self-reported alcohol intake in surveys usually covers only half the amount sold (2, 3). This may be explained partly by sampling errors, i.e., heavy drinkers may be

less likely to participate in surveys. In addition, the discrepancy may be explained by response errors, e.g., difficulties in recall of drinking practices and culturally determined socially desirable answers. Underreporting of alcohol intake by individuals is common for all available methods, which does not, however, necessarily mean these methods are not of value. When each individual underreports a similar proportion, the *level* of alcohol intake is underestimated, but *ranking* is correct (figure 1). However, when underreporting is non-proportional and different between subjects, ranking is also affected. The relation between alcohol intake and health may then be weakened.

In this paper, we systematically review the literature on alcohol intake assessment methods, to establish which method could be able to set limits. In contrast to previous reviews (4, 5), we quantify the performance of the methods and distinguish the *ranking* of individuals from estimating the *level* of intake.

MATERIALS AND METHODS

Literature search

Papers published after 1984 were obtained from MEDLINE and the ALCDOD database from the Netherlands Alcohol Documentation Centre. Search items used were: alcohol, drinking, or alcoholic beverages in combination with questionnaire, validity, repro-

Received for publication May 12, 1997, and accepted for publication November 4, 1998.

From the Division of Human Nutrition and Epidemiology, Wageningen Agricultural University, Wageningen, The Netherlands.

Reprint requests to Dr. Gerda I. J. Feunekes, Unilever Research, Consumer Science Unit, P.O. Box 114, 3130 AC Vlaardingen, The Netherlands.

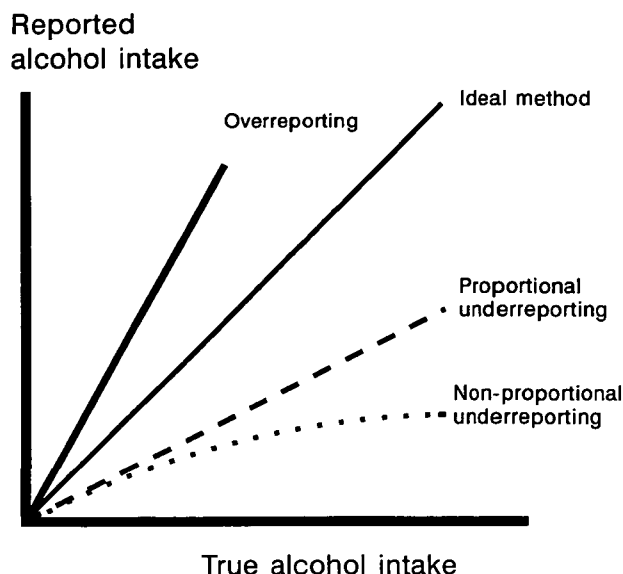


FIGURE 1. Alcohol intake reporting: hypothetical curves of reported alcohol intake compared with true intake.

ducibility, reliability, recall, memory, misclassification, overreporting, or underreporting. We only included papers in English, which comprised 98 percent of the obtained references. Literature on alcohol abuse, alcohol dependency, and drunk driving was excluded, as were studies with alcoholics. Additional references were derived from citations. Because most of the evidence on alcohol and health has come from studies that have used food frequency questionnaires, we included reports on the relative validity or reproducibility of food frequency questionnaires used in major studies (6–11).

Inclusion criteria

Of 84 papers identified in our literature search, 33 (39 percent) were relevant to our review objectives. We excluded studies that assessed alcohol intake in the distant past or post-mortem ($n = 11$). Because of our interest in the general population, studies in special samples were excluded, i.e., studies in students ($n = 15$), pregnant women ($n = 6$), patients ($n = 4$), prisoners ($n = 1$), and subjects at a North Pole army post ($n = 1$). Thirteen additional papers were excluded for one of the following reasons: insufficient description of the type of method, inconsistency in procedures, no within-subjects design, focus on frequency of heavy drinking rather than total intake, or another paper from the data set was already included.

Data collection and analysis

From the selected studies, for *all* methods applied in these studies, we abstracted: length of reference

period, i.e., number of days, weeks, or months, or the ambiguous “usual intake”; beverage specificity, i.e., beer, wine, and spirits separately versus only the total number of alcoholic drinks, and administration mode, i.e., face-to-face interview, telephone interview, self-administered questionnaire, and diary. For each method, mean, standard deviation, number of subjects, and percent of nonusers were abstracted. In addition, Pearson or Spearman coefficients of correlation were obtained. Several studies presented the results for subgroups only, so sometimes two or more data points could be derived from one study. In the results, we therefore refer to “observations” instead of “studies.” The performance of the methods was evaluated quantitatively. We addressed both the assessment of the level of intake and the ranking of individuals according to intake. In all analyses, results of individual observations were weighed by the inverse of the squared standard error of the reported value.

To assess the determinants of the variation in level of intake, we examined the type of method used, length of reference period, beverage specificity, and administration mode. Analysis of variance (SAS 6.11, SAS Institute Inc., Cary, North Carolina) was carried out on all eligible studies which assessed validity of a method and reported mean intake, sample size (n) and a measure of sampling variability, i.e., standard error or standard error of the mean. We adjusted for “study” to account for between-study differences such as sex ratio, age, drinking culture, and social acceptability of drinking. In addition to the quantitative analysis, we carried out a qualitative evaluation to be able to use all studies which fulfilled the selection criteria, including studies that did not have all data required for the quantitative analysis. We checked whether the significant differences in the quantitative analysis were in line with the results of all 33 papers. When the results differed less than 5 percent, the results were considered to be in agreement. For example, if method X appeared to yield a 10 percent higher estimate than method Y in the quantitative analyses, each paper which made a comparison between X and Y was checked for agreement, and a 5–15 percent higher estimation by X would count as “in agreement.”

To describe ranking, data were analyzed for men and women, and age groups separately.

RESULTS

Of 33 eligible papers, 30 assessed relative validity of alcohol intake assessment methods by comparing the estimates of alcohol intake from up to five methods (2, 6–34), but only 12 papers had the quantitative data as

TABLE 1. Characteristics of 30 validity studies included in the review, which all assessed the level of alcohol intake with two or more methods

Author, publication year, and country (ref. no.)	Sample size (males/females)	Methods*
<i>12 studies included in the quantitative analyses (analyses of variance), comprising 39 observations (total of 15,028 subjects)</i>		
Ocké et al., 1996, Netherlands (11)	63/58	FFQ ^{bs} , 24-hour recalls ^{bs}
Goldbohm et al., 1994, Netherlands (10)	59/48	FFQ ^{bs} , PD ^{bs}
Goransson et al., 1994, Sweden (16)	930	QF ^{bs} , extended QF ^{bs}
Single and Wortley, 1994, Canada (3)	7,702	QF, RD
Gavaler and Love, 1992, United States (21)	-/128	QF ^{bs} , PD ^{bs}
Munger et al., 1992, United States (8)	-/44	FFQ ^{bs} , 24-hour recalls ^{bs}
Giovannucci et al., 1991, United States (7)	136/173	QF ^{bs} , PD ^{bs}
Russell et al., 1991, United States (23)	4,367	QF, QF ^{bs}
Flegal, 1990, United States (25)	107/121	QF ^{bs} , PD ^{bs}
Midanik et al., 1989, United States (27)	275/260	QF ^{bs} , RD ^{bs}
Lemmens et al., 1988, Netherlands (28)	399	PD ^{bs} , RD
Pietinen et al., 1988, Finland (6)	158/-	FFQ ^{bs} , PD ^{bs}
<i>Studies (k = 18, and altogether 36,658 subjects) that lacked data to be evaluated in the quantitative analyses, and which were evaluated qualitatively instead</i>		
Breslin et al., 1995, Canada (12)	7/33	QF ^{bs} , PD ^{bs}
Perrine et al., 1995, United States (13)	30/-	RD ^{bs} , PD ^{bs}
Romelsjo et al., 1995, Sweden (14)	1,054/1,359	QF ^{bs} , extended QF ^{bs}
Searles et al., 1995, United States (15)	151/-	QF, 24-hour recalls ^{bs}
King, 1994, United States (17)	237	F ^{bs} , FFQ ^{bs}
Midanik, 1994, United States (18)	2,058	QF, extended QF
Williams et al., 1994, United States (19)	11,208/9,888	QF, extended QF ^{bs}
Wyllie et al., 1994, New Zealand (20)	1,528	QF, extended QF, RD
Feskanich et al., 1993, United States (9)	127/-	FFQ ^{bs} , PD ^{bs}
Lemmens et al., 1992, Netherlands (28)	918	Q, QF, extended QF, PD ^{bs} , RD ^{bs}
Corti et al., 1990, Australia (24)	-/1,356	PD ^{bs} , RD
Hilton, 1989, United States (26)	83	QF ^{bs} , extended QF, PD ^{bs}
Simpura, 1988, Finland (29)	2,907	QF, RD
Sobell et al., 1988, United States (30)	31/31	QF ^{bs} , RD ^{bs}
Fitzgerald and Mulford, 1987, United States (31)	997	QF, extended QF
Redman et al., 1987, Australia (32)	778	QF, PD
Hilton, 1986, United States (33)	1,772	Q (situation-specific) ^{bs} , extended QF ^{bs}
Williams et al., 1985, United States (34)	105	Extended QF ^{bs} , PD ^{bs}

* QF, quantity frequency; F, frequency; Q, quantity; FFQ, QF within food frequency questionnaire; PD, prospective diary; RD, retrospective diary; 24-hour recall, series of 24-hour recalls; ^{bs}, beverage-specific questions.

required (table 1). Eleven of 33 papers had quantitative data of test-retest reproducibility of alcohol intake estimation (6–8, 11, 18, 19, 22, 27, 34–36). The 12 validation studies yielded 39 data points to evaluate the level estimation, whereas 18 studies lacked the data required and could be evaluated qualitatively only. For the evaluation of ranking, 30 correlations from 12 validation studies, and 29 correlations from 11 reproducibility studies were available.

From all methods evaluated in the papers, we identified five main types of methods. The “quantity frequency method” (QF) is a simple method that is often used to assess alcohol intake in a specified period. It consists of one question to assess the average frequency (“How often do you drink?”) combined with a question on the average quantity (“How many drinks per occasion?”).

The quantity frequency method is sometimes extended with questions on variability of drinking practices (e.g., week/weekend, binges), and/or location-specific consumption (e.g., amount at home, in bar). The “graduated frequency method” estimates frequencies over the full range of quantities consumed, starting with the highest amount consumed, and consecutively asking frequencies for all smaller amounts. We consider all these extensions of the QF will be in the category “extended QF”; their numbers were too small to enable separate analyses. Instead of summarizing the drinking habits of subjects for a longer period, as done in the QF, subjects can be asked to record or recall alcohol intake for specific days. Recording intake in a diary can be done either prospectively, in a “prospective diary” (PD), or retrospectively. The “retrospective diary” (RD), in which we

incorporated the "time-line follow back method," asks subjects to recall their day-by-day intake, with or without help of cognitive cues such as a calendar. Another method to recall intake is a series of (random) recalls of consumption in the past 24 hours ("24-hour recalls") (table 1).

The studies with quantitative data came from relatively few areas—North America, and North and West Europe. Alcohol intake was assessed for a period ranging from 3 days to one year, with some methods referring to "usual" intake without specification. All 24-hour recalls and prospective diaries were beverage specific.

Table 2 summarizes the capability of methods to explain variance in the *level* of alcohol intake and shows substantial variation according to method characteristics. Beverage specificity was the only significant predictor of alcohol intake; when subjects were asked separately about their intake of wine, beer, and spirits, this resulted in a 19 percent higher reported intake. Neither the type of method, the length of the reference period, nor the administration mode were statistically significantly related to reported alcohol intake, although these differences possibly would have

TABLE 2. Reported mean alcohol intake (drinks/week), by type of method, beverage specificity, length of reference period, and mode of administration, based on 39 observations from 12 studies with a total of 15,028 subjects (see table 1 for studies included)

	No. of observations per level	Mean alcohol intake \pm SE*
Type of method*		
Quantity frequency (QF)	21	6.2 \pm 1.5
Retrospective diary (RD)	4	5.1 \pm 1.8
Prospective diary (PD)	11	6.2 \pm 1.6
Series of 24-hour recalls (24-hour recalls)	3	7.5 \pm 13.8
Beverage-specific*		
No	8	5.4 \pm 0.5†
Yes	30	6.4 \pm 0.5
Reference period*		
3–7 days	7	5.7 \pm 2.2
8–14 days	4‡	7.7 \pm 7.8
15–30 days	8	6.2 \pm 3.9
One year	16	6.0 \pm 2.0
"Usual"	4	6.0 \pm 2.2
Administration mode*		
Face-to-face interview	5§	3.1 \pm 7.1
Self-administered questionnaire	24	6.1 \pm 4.9
Diary	4	6.1 \pm 4.9
Telephone interview	6¶	8.6 \pm 20.9

* Adjusted for "study." SE, standard error.

† Significantly different from non-beverage specific ($p < 0.001$); $n = 38$.

‡ Two 24-hour recalls, two PD.

§ Three RD, two PD.

¶ Four QF, one RD, one series of 24-hour recalls.

reached significance if individual data had been available for analysis.

As indicated in figure 1, underreporting might be proportional or non-proportional to the level of intake. To study this, we analyzed the data separately for the 19 and 20 observations of groups, with the lowest and highest intake according to the commonly used quantity frequency method: ≤ 7 drinks/week (average 4 (standard deviation (SD) 3) drinks/week) compared with > 7 drinks/week (average 10 (SD 1) drinks/week). For both levels of intake, the patterns were similar, but the differences between methods were larger, suggesting that these systematic reporting errors tended to be proportional to the level of intake (figure 2). The type of method appeared to explain additional variance when combined with beverage specificity in the same model; retrospective diaries gave intakes 20–22 percent below the intake assessed with the quantity frequency method or prospective diaries. No independent effects were detected from length of reference period, and administration mode.

In the additional qualitative evaluation of all 30 studies, five studies appeared to have findings that were in agreement (within 5 percent) with the difference between beverage-specific and non-beverage-specific methods, and two were not. The remaining 23 studies did not include required data to examine beverage specificity. The differences between methods were weakly supported by the qualitative evaluation: in five studies, the findings were in agreement (within 5 percent); six studies were not in agreement; and 11 studies missed the data required for this comparison.

Table 3 shows the ability of the alcohol intake assessment methods to rank individuals. Association between alcohol intake assessed with different methods was reasonable, with validity correlations in different studies ranging from 0.32 to 0.90, and the weighted averages between 0.63 and 0.73. Thus, ranking capability for alcohol intake appeared to be better than for nutrient or energy intake (6, 10, 11). The diagonal in table 3 shows weighted averages of test-retest correlations ranging from 0.84 to 0.88, based on correlations that ranged from 0.75 to 0.99 in the separate studies (6–8, 11, 18, 19, 22, 27, 34–36). Thus, test-retest correlations clearly exceeded validity correlations.

DISCUSSION

From our analysis of the literature, the sober facts on assessment of alcohol intake are that retrospective diaries tend to give intake estimates that are about 20 percent lower than quantity frequency and prospective diaries. Additional assessment of the type of alcoholic drinks gave roughly a 20 percent higher estimated

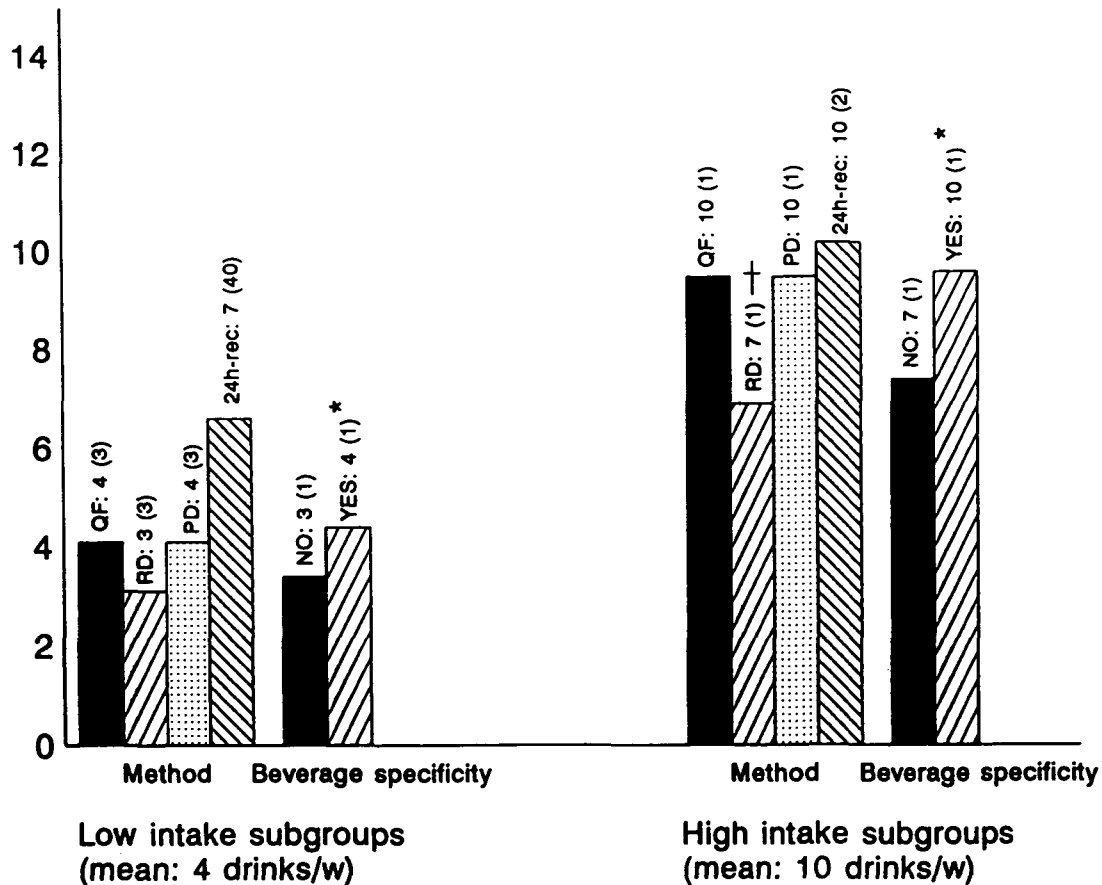


FIGURE 2. Average alcohol intake in drinks/week (standard error of the mean) assessed by four types of methods[‡] and by beverage specificity; separately for subgroups with low and high alcohol intake according to the quantity frequency method. *Beverage specificity: $p < 0.001$ in low intake subgroups and $p = 0.07$ in high intake subgroups. †Retrospective diary gives lower alcohol intake than quantity frequency or prospective diary ($p < 0.01$). ‡QF, quantity frequency; RD, retrospective diary; PD, prospective diary; 24h-rec, 24-hour recalls.

intake. The length of reference period and the administration mode of the questionnaire was not related to the reported level of alcohol intake. In contrast to the systematic differences in the level of alcohol intake, the ranking capacity was satisfactory for most methods evaluated. So, epidemiologists can validly identify associations between alcohol, health, and disease, but this does not directly lead to clarity on sensible limits for acceptable alcohol intake.

Validity of the quantitative approach

Regarding the quality of the studies and the data underlying our analyses, it should be realized that many methodological papers were not relevant to our purposes because of non-representative groups and/or inappropriate design. Even of the 33 eligible papers, only 12 provided the basic quantitative and methodological data required to address level of intake, and 21 studies provided data relevant to ranking. However,

results of qualitative analyses were generally in line with our main findings.

Because we had no access to the original data, we had to rely on published data with population means and correlations. Systematic differences between methods could have been assessed more accurately if we could have compared reported alcohol intake at the individual instead of the aggregate level. Further, a very limited number of papers fulfilled inclusion criteria and had basic quantitative data available. Wine drinking countries in Southern Europe had no methodological studies available, which limited cultural variability in the data. Therefore, we could not evaluate the role of cultural differences in drinking habits, e.g., timing (at daily meal, at celebrations) and situation (alone, or with family, friends, or colleagues). Furthermore, several potentially relevant factors have been reported to affect alcohol intake, but the available data did not allow us to make allowance for these factors in the quantitative analysis. For example, a high number and

TABLE 3. Ranking of individuals according to alcohol intake as reported in various studies: weighted averages of test-retest and validity correlations* of currently available methods

Method	QF	Extended QF	RD	PD	24-hour recalls
QF (quantity frequency)	0.88 (0.75, 0.99)† n‡ = 1,211				
Extended QF (QF with questions on variability or location-specific drinking)	0.63 (0.59, 0.90) n = 24,045	0.88 (0.83, 0.98) n = 317			
RD (retrospective diary)	0.67 (0.66, 0.74) n = 5,855	0.66 (0.66) n = 918	—§		
PD (prospective diary)	0.71 (0.61, 0.90) n = 1,722	0.73 (0.57, 0.89) n = 2,002	0.65 (0.51, 0.65) n = 948	0.84 (0.84) n = 399	
24-hour recalls (series of 24-hr recalls)	0.68 (0.32, 0.90) n = 165	—	—	—	—

* Test-retest correlations are on the diagonal axis, all others are between-method correlations.

† Minimum and maximum values as reported in the studies are shown in parentheses.

‡ n, total number of subjects included.

§ —, no data available.

range of answering categories (37) and assessment of glass/container size (vs. standard size) (38) may increase reported alcohol intake. However, we realize that a general phenomenon in food intake research is that “the more you ask, the more overreporting you get” and this may also apply to alcohol intake assessment. Because no validated independent marker of alcohol intake is currently available, we followed the suggestion of Midanik (4) and concentrated on the relative validity of the methods. The question “What is the best alcohol assessment method?” has to be answered separately for the main objectives of the alcohol assessment. For the purposes of epidemiologic research, we need particularly a valid ranking, and to advise the population we need in addition a valid estimation of level.

Validity of ranking

Despite the difficulties in assessing level of intake, ranking was satisfactory, as exemplified by the high validity and test-retest correlations. Because such correlations were of the same order of magnitude and tend to be similar in many studies (10), Spearman and Pearson correlation coefficients were combined in the analyses. In theory, these high correlations might result from consistent under- and overreporters, leading to so-called correlated errors. However, Goldbohm et al. (10) have published data which showed that 78 percent of the variance was due to between-subject

variance, 12 percent to random within-subject variation, and only 10 percent to correlated within-subject variance. This is much better than for other dietary factors, and explains why epidemiologic studies have been able to find associations between alcohol and health.

For epidemiologic studies, most methods will suffice as they all rank well. In this case, it is important to account for variability in drinking (e.g., day-to-day variation and binge drinking). Simple methods do not rank in a poorer manner than more elaborate methods, so the cheaper method of quantity frequency assessment may be favored.

Validity of level

Our results differ from those of previous reviews (4, 22) in which the quantity frequency method generally gave lower estimations of intake. However, this might be due to the fact that the papers which were suitable for our quantitative analysis were different from the type of papers usually considered in alcohol intake assessment reviews. A third of the papers came from epidemiologic questionnaires which assessed the composition of the diet, thereby including alcohol. Because most findings regarding the relation between alcohol and health are derived from such questionnaires, these papers yielded important information. Possibly the context of the alcohol questions affects the outcome to a larger extent than expected.

Improving methods

Methods used to evaluate new assessment techniques can be improved. First, with respect to questionnaire development in different cultures or circumstances, protocol analysis in which people think out loud about the questions asked (39) is a worthwhile exercise to ensure that questions are clear and that the wording is optimal. Second, concerning study design, cross-over designs have not often been used in studies which validate alcohol intake assessment methods. Assessment of alcohol intake with an extensive method may affect alcohol intake assessed with a shorter method afterwards. Split half design has been used to exclude such a carry-over effect (40). Predictive validity has been assessed with the use of several methods in an epidemiologic study (41) to compare their risk estimates. Last, research in this area would benefit from a good marker of alcohol intake in the "low to moderate" intake range. Although such markers are not yet available, we might be able to work creatively with repeated assessment of short-term intake markers such as saliva, sweat, and breath (15). The use of markers could also help to determine whether underreporting of alcohol intake is also proportional to the actual intake at very low and at very high levels, exceeding the socially acceptable range. Until now, biologic markers have been "validated" against self-reported intake. Possible biologic markers should ideally be tested in a controlled study in which subjects consume a specific amount of alcohol during a longer period.

Conclusion

To advise people, and to define "sensible limits," methods to assess alcohol intake need to give a good estimation of the level of intake. Ranking alone is not sufficient. Considering the methods that could be evaluated in this review, the simplest methods appear to be very much neglected. For instance, the "usual frequency" method, which is often used in surveys, has been evaluated only once (22). In general, we would like to emphasize that ex-drinkers and lifetime abstainers should be marked as different groups. Additionally, the reference period should be chosen which fits the goal: a study on traffic accident and drinking needs another reference period than alcohol intake and liver diseases.

Although our data suggest that the quantity frequency method or prospective diary yield the highest levels of estimated intake, independent methods and/or biologic markers are needed to show which methods are closer to true intake. So, only when there is sufficient evidence that alcohol intake is underesti-

mated in a population, we recommend the assessment of both the quantity and frequency component of alcohol consumption and that these components be assessed separately for all types of alcoholic beverages.

ACKNOWLEDGMENTS

This review was supported by a grant from the European Branch of the International Life Sciences Institute (ILSI).

REFERENCES

1. Inter-Departmental Working Group, Department of Health. Sensible drinking. London: Inter-Departmental Working Group. Department of Health, 1995.
2. Høyer G, Nilssen O, Brenn T, et al. The Svalbard study 1988-89: a unique setting for validation of self-reported alcohol consumption. *Addiction* 1995;90:539-44.
3. Single E, Wortley S. A comparison of alternative measures of alcohol consumption in the Canadian National Survey of alcohol and drug use. *Addiction* 1994;89:395-9.
4. Midanik LT. The validity of self-reported alcohol use: a literature review and assessment. *Br J Addict* 1988;83:1019-29.
5. Litten RZ, Allen JP, eds. Measuring alcohol consumption: psychosocial and biochemical methods. Totowa, NJ: Humana Press, 1992.
6. Pietinen P, Hartman AM, Haapa E, et al. Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. *Am J Epidemiol* 1988;128:655-66.
7. Giovannucci E, Colditz G, Stampfer MJ, et al. The assessment of alcohol consumption by a simple self-administered questionnaire. *Am J Epidemiol* 1991;133:810-17.
8. Munger RG, Folsom AR, Kushi LH, et al. Dietary assessment of older Iowa women with a food frequency questionnaire: nutrient intake, reproducibility, and comparison with 24-hour dietary recall interviews. *Am J Epidemiol* 1992;136:192-200.
9. Feskanich D, Rimm EB, Giovannucci EL, et al. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J Am Diet Assoc* 1993;93:790-6.
10. Goldbohm RA, van den Brandt PA, Brants HAM, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr* 1994;48:253-5.
11. Ocké MC, Bueno-de-Mesquita HB, Pols MA, et al. The Dutch EPIC Food Frequency Questionnaire. II. Relative validity and reproducibility for nutrients. *Int J Epidemiol* 1997;26(Suppl 1):S49-S58.
12. Breslin FC, Okeeffe MK, Burrell L, et al. A comparison of two report formats for alcohol and caffeine use. *Psychology Health* 1995;10:439-48.
13. Perrine MW, Mundt JC, Searles JS, et al. Validation of daily self-reported alcohol consumption using interactive voice response (IVR) technology. *J Stud Alcohol* 1995;56:487-90.
14. Romelsjö A, Leifman H, Nyström S. A comparative study of two methods for the measurement of alcohol consumption in the general population. *Int J Epidemiol* 1995;24:929-36.
15. Searles JS, Perrine MW, Mundt JC, et al. Self-report of drinking using touch-tone telephone: extending the limits of reliable daily contact. *J Stud Alcohol* 1995;56:375-82.
16. Göransson M, Hanson BS. How much can data on days with

- heavy drinking decrease the underestimation of true alcohol consumption? *J Stud Alcohol* 1994;55:695-700.
17. King AC. Enhancing the self-report of alcohol consumption in the community: two questionnaire formats. *Am J Public Health* 1994;84:294-6.
 18. Midanik LT. Comparing usual quantity/frequency and graduated frequency scales to assess yearly alcohol consumption: results from the 1990 US National Alcohol Survey. *Addiction* 1994;89:407-12.
 19. Williams GD, Proudfit AH, Quinn EA, et al. Variations in quantity-frequency measures of alcohol consumption from a general population survey. *Addiction* 1994;89:413-20.
 20. Wyllie A, Zhang J-F, Casswell S. Comparison of six alcohol consumption measures from survey data. *Addiction* 1994;89:425-30.
 21. Gavalier JS, Love G. Detection of the relationship between moderate alcoholic beverage consumption and serum levels of estradiol in normal postmenopausal women: effects of alcohol consumption quantification methods and sample size adequacy. *J Stud Alcohol* 1992;53:389-94.
 22. Lemmens P, Tan ES, Knibbe RA. Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. *J Stud Alcohol* 1992;53:476-86.
 23. Russell M, Welte JW, Barnes GM. Quantity-frequency measures of alcohol consumption: beverage-specific vs global questions. *Br J Addict* 1991;86:409-17.
 24. Corti B, Binns CW, Howat PA, et al. Comparison of 7-day retrospective and prospective alcohol consumption diaries in a female population in Perth, Western Australia—methodological issues. *Br J Addict* 1990;85:379-88.
 25. Flegal KM. Agreement between two dietary methods in the measurement of alcohol consumption. *J Stud Alcohol* 1990;51:408-14.
 26. Hilton ME. A comparison of a prospective diary and two summary recall techniques for recording alcohol consumption. *Br J Addict* 1989;84:1085-92.
 27. Midanik LT, Klatsky AL, Armstrong MA. A comparison of a 7-day recall with two summary measures of alcohol use. *Drug Alcohol Depend* 1989;24:127-34.
 28. Lemmens P, Knibbe RA, Tan F. Weekly recall and diary estimates of alcohol consumption in a general population survey. *J Stud Alcohol* 1988;49:131-35.
 29. Simpura J. Comparison of indices of alcohol consumption in the Finnish 1984 drinking habits survey data. *Drinking Drug Pract Surv* 1988;22:3-10.
 30. Sobell LC, Sobell MB, Leo GI, et al. Reliability of a timeline method: assessing normal drinkers' reports of recent drinking and a comparative evaluation across several populations. *Br J Addict* 1988;83:393-402.
 31. Fitzgerald JL, Mulford HA. Self-report validity issues. *J Stud Alcohol* 1987;48:207-11.
 32. Redman S, Sanson-Fisher RW, Wilkinson C, et al. Agreement between two measures of alcohol consumption. *J Stud Alcohol* 1987;48:104-8.
 33. Hilton ME. Inconsistent responses to questions about alcohol consumption in specified settings. *Am J Drug Alcohol Abuse* 1986;12:403-13.
 34. Williams GD, Aitken SS, Malin H. Reliability of self-reported alcohol consumption in a general population survey. *J Stud Alcohol* 1985;46:223-7.
 35. Longnecker MP, Newcomb PA, Mittendorf R, et al. The reliability of self-reported alcohol consumption in the remote past. *Epidemiology* 1992;3:535-9.
 36. Webb GR, Redman S, Gibberd RW, et al. The reliability and stability of a quantity-frequency method and a diary method of measuring alcohol consumption. *Drug Alcohol Depend* 1991;27:223-31.
 37. Hughes SP, Dodder RA. Alcohol consumption indices: format comparisons. *J Stud Alcohol* 1988;49:100-3.
 38. Lemmens P. The alcohol content of self-report and "standard" drinks. *Addiction* 1994;89:593-602.
 39. Midanik LT, Hines AM. The alcohol content of self-report and "standard drinks." "Unstandard" ways of answering standard questions: protocol analysis in alcohol survey research. *Drug Alcohol Depend* 1991;27:245-52.
 40. Wagenaar AC, Komro KA, McGovern P, et al. Effect of a saliva test pipeline procedure on adolescents self-reported alcohol use. *Addiction* 1993;88:199-208.
 41. Laforge RG, Williams GD, Parker DA, et al. Alcoholic beverage type, recall period effects and functional disability: evidence from the 1983 NHIS. *Drug Alcohol Depend* 1990;25:257-72.