British Journal of Nutrition (2002), **88**, 427–434
DOI: 10.1079/BJN2002671

Concurrent validity of alcohol consumption measurement in a 'healthy' population; quantity-frequency questionnaire *v.* dietary history interview

Lando L. J. Koppes[1], Jos W. R. Twisk[1], Jan Snel[2] and Han C. G. Kemper[1]*

[1]*Institute for Research in Extramural Medicine, VU University Medical Center, Amsterdam, The Netherlands*
[2]*Department of Psychology, Faculty of Social and Behavioural Sciences, University of Amsterdam, Amsterdam, The Netherlands*

(Received 23 August 2001 – Revised 21 March 2002 – Accepted 18 May 2002)

Self-reports of alcohol consumption account for approximately 50 % of the reported sales of alcohol. In the absence of a gold standard, it is not known how accurately different methods of measurement reflect actual consumption and whether under-reporting varies among different populations. The objective of the present study was to compare the consumption reported by the widely used quantity-frequency questionnaire (QFQ) with that reported in a cross-check dietary history interview (DHI), which has higher face validity. In 171 male and 197 female subjects of the Amsterdam Growth and Health Longitudinal Study (mean age 36 years), alcohol consumption was assessed by both the QFQ and the DHI. Most subjects reported a moderate consumption of alcohol by both measures. Spearman correlation coefficients were high (0·77 and 0·87 in men and women respectively). Overall, greater alcohol consumption was reported using the DHI. The difference between the DHI and QFQ reports was usually greater for wine than for beer. Backward stepwise regression analysis showed that the difference in reporting was positively related to a more irregular drinking pattern, and in wine drinkers to the square of the QFQ report. Sex, drinking alone or with others and the CAGE (acronym for four questions on drinking behaviour) score were not related to the difference in reporting. The precision of DHI estimation from QFQ reports and other factors was low. Serious questions arise as to the validity and precision of alcohol consumption measurements based on the QFQ alone. QFQ information may be improved by incorporating questions on the type of beverage and drinking patterns.

Alcohol consumption: Measurement: Validity: Modification

The extent to which self-reported measurements of alcohol consumption agree with the amount of alcohol that has been sold differs widely (Midanik, 1982; Redman *et al.* 1987; Lemmens *et al.* 1992; Single & Wortley, 1994; Wyllie *et al.* 1994; Romelsjö *et al.* 1995). This divergence in the reporting of alcohol consumption threatens the validity and comparability of alcohol research, because under-reporting has dramatic effects on the magnitude of the resulting regression coefficients and cut-off values used. If, hypothetically, 50 % of the alcohol consumed is systematically not reported, the relationships found with other factors are twice the actual values. Ranges of alcohol intake (and cut-off values) are also dramatically affected by under-reporting. For example, if in a certain study a mortality risk was found to be lowest in subjects who drink up to 3·0 units/d and 50 % of the alcohol consumed was not reported, then the lower mortality risk would, in fact, range up to 6·0 units/d. The use of different techniques to measure alcohol consumption can result in different estimates that lead to different conclusions (Rehm *et al.* 1999).

If the relative magnitude of the under-reporting of various measurements of alcohol consumption was known, estimates of the levels of consumption and estimates of the relationship of alcohol with other variables could be corrected. This would make it possible to compare the results from studies using different methods to measure alcohol consumption. However, the under-reporting of alcohol consumption may be related to more factors than just the method of measurement. For example, the level of under-reporting may differ between men and women,

Abbreviations: AGAHLS, Amsterdam Growth And Health Longitudinal Study; CAGE, acronym for four questions on drinking behaviour (for details, see p. 429); DHI, dietary history interview; QFQ, quantity-frequency questionnaire.

* **Corresponding author:** Professor Dr Han C. G. Kemper, fax +31 20 444 8181, email HCG.KEMPER.EMGO@MED.VU.NL

it may increase with the level of consumption, and it may differ depending on the type of beverage consumed (Midanik, 1982; Fitzgerald & Mulford, 1987; Kühlhorn & Leifman, 1993; Romelsjö *et al.* 1995; Rehm *et al.* 1999).

In the present study, the amount of alcohol reported in the widely used quantity-frequency questionnaire (QFQ) is compared with the amount of alcohol reported in a cross-check dietary history interview (DHI). The DHI has a higher face validity than the QFQ for the following reasons: (1) the DHI incorporates extensive indications with regard to drinking locations, situations, type of beverage and unit size; (2) the DHI is subject to socially desirable answers with regard to alcohol consumption to a lesser degree; (3) the focus of the DHI is on total dietary intake; (4) the DHI includes questions with which previous answers are cross-checked. It is therefore expected that the results of the present study will provide an indication of the precision and level of misreporting in the QFQ. Furthermore, the present study investigates whether the difference between the level of alcohol consumption reported in the DHI and in the QFQ is related to the consumption reported in the QFQ, and whether the difference is related to sex, irregularity of drinking pattern, drinking alone or with others, beverage of preference, or with the CAGE (acronym for four questions on drinking behaviour; for details, see p. 429) score.

Methods

Subjects

This study is part of the Amsterdam Growth and Health Longitudinal Study (AGAHLS; Kemper, 1985, 1995). The AGAHLS started in 1977 to measure the lifestyle, health and psychological characteristics of almost 600 13-year-old boys and girls. The present analyses are based on data from the men (*n* 171) and women (*n* 197) (mean age 36·1 (SD 0·7) years) who attended the ninth follow-up measurement that was carried out in 2000. No baseline differences in alcohol consumption were found between subjects who dropped out and those who attended all measurements (Koppes *et al.* 2000).

Measurement of alcohol consumption

The subjects visited the AGAHLS laboratory for 1 d. Various measurements were performed, including two to assess weekly alcohol consumption: a cross-check DHI, which was specifically designed for the AGAHLS (Post, 1989), and a QFQ. The order in which the measurements were performed differed for each subject.

The QFQ consisted of two questions only. The first was: 'How often do you consume one or more alcoholic drinks?'. Response options for this first question were: never, monthly or less often, two to four times per month, two to three times per week, or four times per week or more often (coded as 0·00, 0·25, 0·75, 2·50 and 5·50 respectively). Then, those who consumed alcohol were asked: 'How many alcoholic drinks (glasses) do you consume on an average day on which you drink?'. For this question the response options were: 1, 2 or 3, 4 or 5, 6–8, or 9 or more (coded as 1·00, 2·50, 4·50, 7·00, and 10·00 respectively). Consumption per week based on the QFQ was calculated by multiplying frequency by quantity. The QFQ does not specify the type of beverage or the size of a unit. In the Netherlands, a standard glass contains 10 g pure alcohol.

In the DHI, a trained interviewer asked the subjects what, how often and how much they ate and drank during the day. The interviewer concurrently entered the answers in a computer using a food assessment program based on Dishes 98 (Mensink *et al.* 1998). The interview referred to the dietary intake of the previous month. Questions were specific with regard to the time of day (e.g. between lunch and dinner) and the mode of preparation, and covered series of items from the entire range of foods and drinks. The interview lasted for approximately 60 min, during which time all kinds of indications were given to include atypical aspects of dietary intake. As for the other products, an indication of the size of the units of alcoholic beverages consumed was obtained by asking the subjects to indicate which of a selection of various sized glasses matched their unit of consumption. The data on alcohol consumption entered into a computer were converted into units of 10 g alcohol (the standard unit size in the Netherlands) consumed per week.

Other aspects of alcohol consumption

A measure of the irregularity of drinking pattern was obtained by dividing the largest number of units consumed on one occasion during the previous month by the average number of units consumed on an average drinking day, resulting in a value of the maximum number of units consumed on one occasion in the previous month:average number of drinks per occasion. Whether the subject usually consumed alcohol alone or together with others was measured on a five-point scale (a higher value indicating more often together with others). For each subject, beer or wine was determined as the beverage of preference on the basis of which beverage was the source of the greatest amount of pure alcohol reported in the DHI. Only nine subjects consumed the greatest amount of alcohol as spirits. These subjects were added to the category of beer or wine preference (coded as 1·00 and 2·00), except for the two subjects who consumed all their alcohol in spirits (coded 1·50). The CAGE questionnaire was used to obtain a measure of problem drinking (Ewing, 1984). CAGE is the acronym for the questions: Have you ever felt you should cut down on your drinking? Have people annoyed you by criticising your drinking? Have you ever felt bad or guilty about your drinking? Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover (eye-opener)? The score is the number of questions answered with 'yes'.

Data analysis

Three methods were used to analyse the agreement between the DHI and the QFQ. First, Spearman correlation coefficients were calculated for men and women separately, and calculated separately with and without the

subjects who reported no alcohol consumption in both measures. Second, percentage agreement was calculated for men and women separately, based on three categories: non-consumers, moderate consumers and greater than moderate consumers. The upper limits for moderate alcohol consumption in men and women were set at 30 and 20 g pure alcohol/d respectively, which is the standard in the Netherlands (Nederlands Economisch Instituut, 1998). The third method was to measure the difference in reporting as a function of the quantity reported in the QFQ, which is calculated as the regression of the difference between the amounts of alcohol reported in the DHI and the QFQ v. the amount reported in the QFQ. The linearity of this difference between the two measures and that reported in the QFQ was investigated by adding the square of the amount reported in the QFQ to the linear model. The difference in reporting for men and women was established by multiplying sex by the amount reported in the QFQ and adding the result to the model. If the interaction term was significant, stratified analyses were performed. Modifications of the difference in reporting by irregularity of the drinking pattern, alone or with other people who were also drinking, beverage of preference and the CAGE score were investigated accordingly. Backward stepwise regression analysis was performed to create a model with factors that are independently related to the difference in reporting. The initial full model consists of the amount reported in the QFQ, sex, irregularity of drinking pattern, drinking alone or with others, beverage of preference and the CAGE score. Regression analyses were based only on data from subjects who reported that they consumed alcohol in both the QFQ and the DHI.

Results

Table 1 shows the Spearman correlation coefficients between the two measures of alcohol consumption. Especially in women, the coefficients were high. The coefficients for the subjects who reported that they consumed alcohol on at least one of the two measures were only 0·04 smaller than those with the zero–zero ties included. Tables 2 and 3 show the number of men and women reporting non-consumption, moderate consumption and greater than moderate consumption in the QFQ and the DHI. The percentage agreement was relatively high: 82·5 % in men and 71·6 % in women. Most AGAHLS subjects were classified as moderate drinkers of alcohol by both measures. Fewer men and women reported non-consumption or greater than moderate consumption in the QFQ than in the DHI. In the QFQ, 85·4 % of the men and 80·7 % of the women were classified as moderate drinkers. One female subject reported no consumption of alcohol in the QFQ, but moderate consumption in the DHI. This subject was pregnant and consumed no alcohol at the time of the assessments. Her report of alcohol consumption >0·0 units/d in the DHI was because the nutritionist who performed the DHI assessment asked this subject to report her usual 'pre-pregnant' level of alcohol consumption. Therefore, the accuracy of the non-drinking report in the QFQ was, in fact, 100 %.

Table 1. Alcohol consumption reported in the quantity-frequency questionnaire and in the dietary history interview by men and women, with and without subjects who reported no alcohol consumption in both measures*

(Spearman correlation coefficients)

	Men	Women
All subjects	0·81	0·91
Non-consumers excluded	0·77	0·87

* For details of subjects and procedures, see p. 428.

In Table 4, it can be seen that for both men and women, the mean amount of alcohol reported per consumer was moderate in both the QFQ and the DHI, although the average amount that was reported differed between the two measurement methods. In the QFQ, the men reported only 63·5 % of the amount they reported in the DHI, while the amount reported by women in the QFQ was only 54·0 % of the amount they reported in the DHI. Table 4 also shows some other characteristics of alcohol consumption. The irregularity of the drinking pattern was greater in men than in women. Most subjects reported the consumption of alcohol in the company of others. The beverage of preference was wine for the great majority of women, but beer was the beverage of preference slightly more often for men. Of the men, 34 % and of the women 19 % answered 'yes' to at least one of the four CAGE questions, but none of the subjects gave a positive answer to all four CAGE questions.

In performing regression diagnostics on the difference in reporting between the QFQ and the DHI, three model outliers were found (Cook's distances). These outliers were the three subjects who scored the highest in the QFQ; drinking at least four times per week and at least 9·0 units per occasion. Data obtained from these subjects were excluded from the regression analyses.

The relationship between the difference in QFQ and DHI reports and the QFQ report was not modified by sex, irregularity of drinking pattern, drinking alone or with others, or the CAGE score ($P > 0·5$). Significant modification of the relationship was found for beverage of preference. Fig. 1a shows the difference in reporting for subjects who consumed the greatest amount of alcohol as beer. It shows that the relatively greater amount of alcohol reported in the DHI decreased slightly with an increasing amount reported in the QFQ (for trend $P=0·07$). Fig. 1b shows the difference in reporting for subjects who obtained the greatest amount of alcohol as wine. The difference between alcohol consumption reports in the DHI and in the QFQ had a significant ($P<0·001$) quadratic relationship with the report in the QFQ. The regression line fitted through the data had a local high at QFQ 17·1 units/week, where the DHI report is, on average, 9·7 units/week greater than the QFQ report. The significance of the quadratic term of the regression fit in Fig. 1b was largely due to the three subjects who reported 38·5 units/week in the QFQ. Omitting these three subjects resulted in a 37 % smaller, and no longer statistically significant, quadratic term.

Table 5 shows the full and reduced model regression coefficients of the difference between reports in the DHI

Table 2. Men (*n*) reporting the consumption of 0·0, >0·0 but ≤21·0, and >21·0 units alcohol/week in the dietary history interview and in the quantity-frequency questionnaire*

		DHI			
		0·0	>0·0 and ≤21·0	>21·0	Total
QFQ	0·0	9			9
	>0·0 and ≤21·0	7	120	19	146
	>21·0		4	12	16
	Total	16	124	31	171

DHI, dietary history interview; QFQ, quantity-frequency questionnaire.
* For details of subjects and procedures, see p. 428.

and in the QFQ, stratified for beverage of preference. In those who preferred beer, backward stepwise regression analysis resulted in only one factor that was independently related to the difference in reporting: the higher the score on irregularity of drinking pattern, the higher the amount reported in the DHI compared with the QFQ. Consequently, the difference between QFQ and DHI reports was not significant in regular pattern beer drinkers (for whom the maximum number of drinks: average number of drinks was <2·0). In those who preferred wine, the irregularity of drinking pattern and the amount reported in the QFQ were independently related to the difference in reporting. In contrast with regular pattern beer drinkers, wine drinkers who have a maximum number of drinks:average number of drinks ratio <2 still reported a significantly higher amount of consumption in the DHI than the QFQ. The standard errors of the estimate of the reduced models were 7·8 and 6·8 units/week respectively. A footnote to Table 5 shows the equations that correspond to the coefficients of the reduced models in Table 5. It also gives an example of the estimation of the DHI report from the model determinants.

Discussion

In the present study, alcohol consumption was measured in a young and 'healthy' population by means of two different methods, the QFQ and the DHI. The correlation coefficients between the two measures were slightly higher than those reported in most other studies (Feunekes *et al.* 1999; Rehm *et al.* 1999). From such relatively high correlation coefficients, and from the rather high percentage agreement between the QFQ and DHI, one may presume that the measures are valid and appropriate for use in epidemiological studies. However, most AGAHLS subjects reported a much higher amount of alcohol consumption in the DHI than the QFQ.

The question of whether this difference in results from the QFQ and from the DHI is due to under-reporting in the QFQ, or to over-reporting in the DHI, or to both, cannot be answered due to the lack of a gold standard. In general, self-reported alcohol consumption accounts for only half the amount of alcohol sold (Pernanen, 1974). Therefore, over-reporting is hardly ever considered to be a plausible option (Midanik, 1982) and the assessment method that results in the highest amount of reported alcohol consumption is expected to be closest to the truth. The fact that more alcohol consumption was reported in the DHI than in the QFQ is therefore in accordance with the assumption outlined on p. 431, i.e. that the DHI provides more valid information on alcohol consumption than the QFQ.

The consumption reported in the QFQ, on average, was only 54·0 % for women and 63·5 % for men of the amount reported in the DHI. This large difference in mean consumption between the two methods was unexpected, because both assessments were performed on the same day. The subjects may have remembered their answers during the first assessment while responding to the second, and may also have tried to give similar answers. Nevertheless, several reasons can be suggested for the difference found in reporting. The wide variation of indications in the DHI may have helped the subjects to remember drinking occasions and amounts. In other studies, higher amounts have also been reported when the questions on alcohol consumption were more extensive (Fitzgerald &

Table 3. Women (*n*) reporting the consumption of 0·0, >0·0 but ≤14·0, and >14·0 units alcohol/week in the dietary history interview and in the quantity-frequency questionnaire*

		DHI			
		0·0	>0·0 and ≤14·0	>14·0	Total
QFQ	0·0	31	1		32
	>0·0 and ≤14·0	21	105	33	159
	>14·0		1	5	6
	Total	52	107	38	197

DHI, dietary history interview; QFQ, quantity-frequency questionnaire.
* For details of subjects and procedures, see p. 428.

Table 4. Characteristics of the subjects in relation to alcohol consumption*

	Men (*n* 171)				Women (*n* 197)			
n	Mean	SD	*n*	%	Mean	SD	*n*	%
Mean consumption (units per week)								
QFQ	9·0	10·3			5·4	5·5		
DHI	14·2	13·9			9·9	9·6		
Drinking pattern irregularity†‡	2·5	1·4			1·8	1·3		
Drinking alone or with others‡								
Always alone			0	0			0	0
Usually alone			5	3			4	2
As often alone as with others			25	15			14	8
Usually with others			58	36			49	30
Always with others			74	46			98	59
Beverage of preference§								
Beer			82	53			15	11
Wine			73	47			127	89
Cage‡								
0			107	66			134	81
1			30	19			17	10
2			20	12			10	6
3			5	3			4	2
4			0	0			0	0

QFQ, quantity-frequency questionnaire; DHI, dietary history interview; CAGE, acronym for four questions on drinking behaviour (for details see p. 429).
* For details of subjects and procedures, see p. 428.
† Drinking pattern irregularity is defined as the maximum number of beverages consumed on one occasion in the previous month divided by the average number of beverages per d on which alcohol is consumed.
‡ Subjects with alcohol consumption by QFQ > 0·0 only.
§ Subjects with alcohol consumption by QFQ > 0·0 and by DHI > 0·0 only.

Mulford, 1987; Single & Wortley, 1994). The abundance of detailed questions may also have made it difficult for subjects to give answers that were as socially desirable as they might have wished. The DHI can be considered as a 'disguised' questionnaire, because no more emphasis was placed on alcohol consumption than on any other beverage or food product. Therefore, the DHI is expected to result in less socially desirable answers. The amount reported in the QFQ may also be smaller than the amount reported in the DHI because unit size is only incorporated in the DHI. However, this may explain only a small part of the difference, as Lemmens (1994) has shown with other results from the Netherlands that the average size of the units that are consumed at home are only 6 % (men) to 12 % (women) larger than standard units. Furthermore, subjects may report modal amounts in the QFQ, whereas the average amount is asked for, thereby failing to report the excess of alcohol consumed on days

Table 5. Variables in the full and reduced models predicting the difference in reporting between the dietary history interview and the quantity-frequency questionnaire for subjects consuming the greatest amount of alcohol in beer or in wine*†‡
(Regression coefficients and 95 % confidence intervals)

	Beer (*n* 94)§				Wine (*n* 200)§			
	Full	95 % CI	Reduced	95 % CI	Full	95 % CI	Reduced	95 % CI
Constant (units/week)	5·8	−5·5, 17·2	−2·2	−5·3, 1·0	3·4	−4·1, 10·9	−2·1	−4·2, −0·0
QFQ (units/week)	−0·2	−0·4, 0·0			0·7	0·3, 1·1	0·9	0·6, 1·3
QFQ^2 ((units/week)^2)					−0·02	−0·03, −0·01	−0·027	−0·038, −0·016
Sex (men as reference)	−2·0	−6·3, 2·2			−0·7	−2·7, 1·3		
Drinking irregularity‖	1·5	0·3, 2·8	2·1	0·9, 3·2	1·5	0·8, 2·1	1·6	0·9, 2·2
Drinking alone or with others	−0·8	−2·7, 1·1			−0·8	−2·1, 0·6		
CAGE	1·3	−0·5, 3·2			1·4	−0·1, 3·0		

QFQ, quantity-frequency questionnaire; CAGE, acronym for four questions on drinking behaviour (for details see p. 429).
* For details of subjects and procedures, see p. 428.
† Equations to estimate DHI, using QFQ and drinking pattern irregularity (Irreg) were as follows: preference for beer: $DHI = QFQ + 2{\cdot}1 \times Irreg - 2{\cdot}2$; preference for wine: $DHI = 1{\cdot}9 \times QFQ - 0{\cdot}027 \times QFQ^2 + 1{\cdot}6 \times Irreg - 2{\cdot}1$. For example, a man or women reports the consumption of a mean value of two alcoholic beverages (usually wine) on five days of the week by QFQ (QFQ score 10·0), with a maximum of eight beverages consumed on one occasion in the previous month (drinking pattern $Irreg = 8/2 = 4$). Thus, the estimated consumption based on the DHI would be 20·6 units/week $(1{\cdot}9 \times 10{\cdot}0 - 0{\cdot}027 \times 100 + 1{\cdot}6 \times 4 - 2{\cdot}1)$. In the present example, the estimated amount reported by the DHI is more than twice the amount reported in the QFQ. Since SE = 6·8, the g 5 % CI of the DHI estimate is 7·2, 34·0 units/week.
‡ Regression coefficients refer to a 1·0 higher report on the independent variables.
§ Participants with QFQ > 0 and DHI > 0 only.
‖ Drinking pattern irregularity is defined as the maximum number of beverages consumed on one occasion in the previous month divided by the average number of beverages per d on which alcohol is consumed.

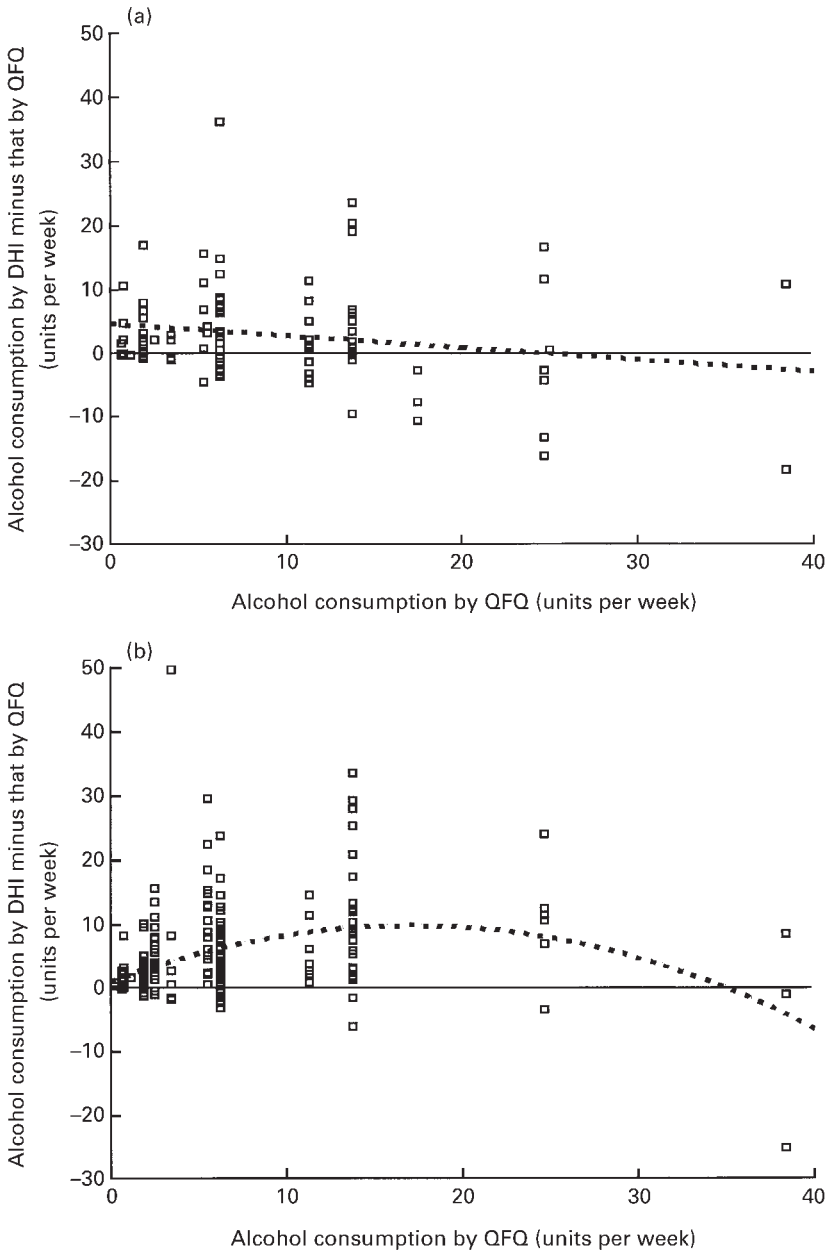

Fig. 1. Scatter plots of the difference in alcohol consumption reported by the dietary history interview (DHI) and the quantity-frequency questionnaire (QFQ) *v.* consumption reported by the QFQ. (a) Subjects who consumed the greatest amount of alcohol as beer (*n* 94); (b) Subjects who consumed the greatest amount of alcohol as wine (*n* 200). —, indicates no difference between QFQ and DHI reports; ▪ ▪ ▪ ▪, indicates regression fit. For details of subjects and procedures, see p. 428.

with unusually high consumption. Finally, the highest response options for the QFQ were ≥4 d per week and ≥9·0 units/d. Coding these as 5·50 and 10·00 respectively may have been too modest. However, post-hoc analyses showed that this might also explain only a minor part of the difference. When the highest response for the QFQ options were recoded to the extremes of 7 d per week and 13·0 units/d, the difference between the DHI and the QFQ reports decreased by 37 % for men and by 16 % for women.

The way in which the difference in reporting between the QFQ and the DHI was related to the QFQ report was not the same for subjects who consumed the greatest amount of alcohol in beer or in wine. Except for those who reported <7·0 units/week in the QFQ, the relative under-reporting in the QFQ appeared to be higher for wine than for beer. Others have also found a general tendency towards greater under-reporting for wine (Midanik, 1982; Fitzgerald & Mulford, 1987; Kühlhorn & Leifman, 1993). This may be related to a greater discrepancy between the home and standard unit size for wine. Moreover, compared with beer-drinking occasions, wine-drinking

occasions (e.g. during dinner) may more easily be forgotten for the QFQ. In the DHI, these occasions are specifically mentioned and incorporated to a greater extent, which might have resulted in relatively more under-reporting of wine in the QFQ. In beer drinkers, the higher amount of alcohol reported in the DHI than in the QFQ decreased slightly with an increase in the amount reported in the QFQ. The difference was absent even in greater-than-moderate drinkers. In subjects who consumed the greatest amount of alcohol in wine, the average greater amount reported in the DHI increased with an increasing amount reported in the QFQ until an amount of 17·1 units/week, when the DHI reporting was estimated to be almost 10·0 units higher. Above a QFQ report of 17·1 units/week, the difference between the reports showed a significant decrease. Thus, in contrast with the findings of Feunekes *et al.* (1999), a linear increase in under-reporting was not found in the present study. In the present study, in accordance with the findings of Redman *et al.* (1987), the subjects with the greatest amounts reported in the QFQ showed a relatively small discrepancy between the two measures of alcohol consumption. For these higher QFQ reports, however, the regression fit of the difference compared with the DHI was less stable, due to the small number of subjects reporting $>17{\cdot}0$ units/week in the QFQ.

Irregularity of the drinking pattern was the only factor (apart from the amount reported in the QFQ) that was independently related to the difference in amount of alcohol reported. Subjects with a large difference between their highest occasional amount and their average drinking amount reported relatively more alcohol consumption in the DHI than in the QFQ. This was in accordance with the findings of others (Fitzgerald & Mulford, 1987; Lemmens *et al.* 1992), and was to be expected, because the DHI gives more indications that could help subjects to incorporate unusually high occasional amounts in their reports.

Sex, drinking alone or with others and the CAGE score were not significantly related to the difference between the QFQ and the DHI reports in the full multiple regression model. Though for the CAGE questionnaire, positive trends with the relative under-reporting in the QFQ were seen for both wine and beer (wine $P=0{\cdot}07$, beer $P=0{\cdot}12$). This finding may be explained by a higher tendency to under-report consumption for socially desirable reasons by subjects who report CAGE problems. Due to the extensive questions in the DHI, these subjects may have been unable to under-report in the DHI to the same extent as in the QFQ.

In the present study, a model was built to obtain a synthetic DHI estimate of alcohol consumption. Unfortunately, this model can only explain a minor part of the individual variation in the difference between DHI and QFQ reports. Therefore, for a subject who reports the consumption of a certain amount of alcohol in the QFQ, and whose preferred beverage and level of irregularity in drinking pattern are known, the precision of the DHI estimate of alcohol consumption is low. The 95 % CI of the DHI estimate for a person who reports the consumption of, for example, 7·0 units/week in the QFQ, would include both non-consumption and greater than moderate consumption. Therefore, general practitioners or researchers, for example, who use the QFQ should bear in mind that the precision of the estimated alcohol consumption is low, but that this precision can be slightly improved by integrating information on the preferred beverage and drinking pattern.

Many methods are used to assess alcohol consumption other than the two described in the present study. Here, the DHI was used because of its characteristics of a gold standard, and the QFQ was chosen for its brevity and frequent use. The precision and validity of the alcohol consumption report in the QFQ appeared to be poor. For example, a 7 d recall method, or a method in which the number of drinks of each of wine, beer and spirits was recalled might have resulted in smaller discrepancies with the DHI, and may have broader applicability. Another option to consider in future studies is to add a sub-study with in-depth alcohol intake history to a brief assessment of alcohol consumption in all subjects. Then, by doing analyses described in the present study, one could arrive at a study-specific synthetic estimate of alcohol consumption without having performed extensive assessments in all subjects.

The most important limitation of the present study is the rather small, 'healthy' and age-specific population, the majority of whom reported that they were moderate drinkers in both measures of consumption. Therefore, it may not be correct to extrapolate the results to other age groups, to patient groups, or to individuals with high levels of alcohol consumption. A second limitation is that, due to the lack of a gold standard, and despite the high face validity of the DHI, it is not known whether the DHI really is a superior method to measure alcohol consumption. From the present findings, however, it can be concluded that serious questions arise as to the validity and precision of measuring alcohol consumption purely on the basis of a QFQ. Therefore, serious questions also arise as to the validity of reported estimates of the relationships between alcohol consumption and, for instance, health outcomes. For example, a stronger relationship with HDL-cholesterol found for wine compared with beer could be caused by greater under-reporting of wine. Alternatively, the relationship found between irregularity in drinking pattern and the level of alcohol consumption reported in a QFQ may be under-estimated solely due to the greater under-reporting by those with a greater discrepancy between the maximum and average amount of alcohol consumed per occasion. In conclusion, the present study shows that information based on quantity-frequency measurements of alcohol consumption should be interpreted with the greatest care, but that the measurements may be improved by integrating information on type of beverage and drinking patterns.

Acknowledgements

The authors wish to thank all the subjects in the AGAHLS for investing their time and efforts in the measurements in 2000. The AGAHLS was financially supported by Dutch Heart Foundation grants 76051–79051, Dutch Prevention

Fund grants 28-189a, 28-1106 and 28-1106-1, Dutch Ministry of Well Being and Public Health grant 90–170, the Dairy Foundation on Nutrition and Health, the Dutch Olympic Committee/Netherlands Sports Federation, Heineken Inc. and the Scientific Board of Smoking and Health.

References

Ewing JA (1984) Detecting alcoholism. The CAGE questionnaire. *Journal of the American Medical Association* **252**, 1905–1907.

Feunekes GIJ, Van 't Veer P, Van Staveren WA & Kok FJ (1999) Alcohol intake assessment: the sober facts. *American Journal of Epidemiology* **150**, 105–112.

Fitzgerald JL & Mulford HA (1987) Self-report validity issues. *Journal of Studies on Alcohol* **48**, 207–211.

Kemper HCG (1985) Growth, health and fitness of teenagers: Longitudinal research in international perspective. *Medicine and Sport Science* **20**, Basel: Karger.

Kemper HCG (1995) *The Amsterdam Growth and Health Study: A Longitudinal Analysis of Health, Fitness, and Lifestyle. HK Sport Science Monograph* **6**. Champaign, IL: Human Kinetics.

Koppes LLJ, Kemper HCG, Post GB, Snel J & Twisk JWR (2000) Development and stability of alcohol consumption from adolescence into adulthood: the Amsterdam Growth and Health Longitudinal Study. *European Addiction Research* **6**, 183–188.

Kühlhorn E & Leifman H (1993) Alcohol surveys with high and low coverage rate: a comparative analysis of survey strategies in the alcohol field. *Journal of Studies on Alcohol* **54**, 542–554.

Lemmens PH (1994) The alcohol content of self-report and 'standard' drinks. *Addiction* **89**, 593–601.

Lemmens P, Tan ES & Knibbe RA (1992) Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. *Journal of Studies on Alcohol* **53**, 476–486.

Mensink GBM, Hermann-Kunz E & Thamm M (1998) Der Ernährungssurvey (The nutrition survey). *Gesundheitswesen* **60**, Suppl. 2, S83–S86.

Midanik L (1982) The validity of self-reported alcohol consumption and alcohol problems: a literature review. *British Journal of Addiction* **77**, 357–382.

Nederlands Economisch Instituut (1998) Alcoholgebruik in beeld; Standaardmeetlat (Alcohol consumption in focus; Standard measuring-staff). Rotterdam: NEI.

Pernanen K (1974) Validity of survey data on alcohol use. In *Research Advances in Alcohol and Drug Problems 1*, pp. 355–374 [RJ Gibbins, editor]. New York: Wiley.

Post GB (1989) Nutrition in adolescence: a longitudinal study in dietary patterns from teenager to adult. Haarlem: De Vrieseborch.

Redman S, Sanson-Fisher RW, Wilkinson C, Fahey PP & Gibberd RW (1987) Agreement between two measures of alcohol consumption. *Journal of Studies on Alcohol* **48**, 104–108.

Rehm J, Greenfield TK, 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.

Romelsjö A, Leifman H & Nyström S (1995) A comparative study of two methods for the measurement of alcohol consumption in the general population. *International Journal of Epidemiology* **24**, 929–936.

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

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