

Validity and reproducibility of two semi-quantitative alcohol frequency questionnaires for the Colombian population

Oscar F Herrán* † and María F Ardila

Centro de Investigaciones Epidemiológicas, Observatorio Epidemiológico de Enfermedades Cardiovasculares, Universidad Industrial de Santander, Bucaramanga, Colombia

Submitted 16 November 2004; Accepted 22 September 2005

Abstract

Objective: To determine the relative validity and reproducibility of two alcohol intake frequency questionnaires (AFQ-A; AFQ-B), designed to classify subjects according to their alcohol intake level, in Bucaramanga, Colombia.

Method: One hundred and nine randomly selected subjects, aged between 20 and 60 years, completed three 30-day semi-quantitative alcohol intake records (30-DR). The AFQs were applied three months after the last 30-DR. AFQ-A contained 53 items; AFQ-B contained five items, with the alcoholic drinks for AFQ-B selected by Max_r. The correlation and agreement between alcohol intake assessed with the AFQs and the 30-DR were obtained using Pearson's correlation coefficient (r), Lin's concordance correlation coefficient (ρ_c), Spearman's rank correlation coefficient (r_s), Bland and Altman's limits of agreement (LOA) and Cohen's weighted kappa statistic (K_w).

Results: The reproducibility of the 30-DR was poor; r_s ranged from 0.33 to 0.41. The reproducibility of the AFQs was higher, with r_s between 0.50 and 0.73. The agreement (K_w) of the 30-DR and the AFQs was 0.40. The lower and upper LOA were between -56.4% and -11.0% . The AFQs and 30-DR were well correlated. Assessment of relative validity between the two methods yielded r values for alcohol between 0.52 and 0.60, which reduced to 0.20–0.29 after energy adjustment.

Conclusions: These AFQs may be useful to rank subjects according to their alcohol intake. The AFQ-B is easy and quick to apply, and is also highly cost-effective.

Keywords
Alcohol consumption
Food-frequency questionnaire
Max_r
Colombia

During the last 50 years, Colombia – like many other countries of the world – has experienced demographic and epidemiological transitions. The changes attributed to these transitions include ageing of the population, changed dietary habits to high intakes of foods rich in fats, sugar and salt, and increased alcohol consumption^{1–3}. The prevalence of alcohol consumption has increased in Colombia during the last 15 years, reaching 86% among the population older than 12 years in 1997, and also the age at which consumption begins has decreased to 9 years old^{4–6}. Despite the need to quantify alcohol intake and establish whether or not it is related to dietary habits, chronic diseases, trauma and social or mental diseases, there are no valid instruments available for this purpose in Colombia.

Even when alcohol is not a nutrient, methods used in the study of diet can be applied in the quantification of alcohol consumption⁷. The research on dietary intake and its relationship with chronic diseases has been carried out with simplified questionnaires that ask about the

dietary pattern before the disease. Alcohol frequency questionnaires (AFQs) have been the most useful tool in epidemiological research^{7–9}, not only for their low cost but also because they result in good classification of subjects based on their intake. Moreover, the level of intake can be measured with more validity¹⁰. However, alcohol consumption is a difficult-to-measure exposure. Due its great between-person and within-person variation, it is necessary to follow it for long periods of time^{7–9}. Given that consumption of alcohol is highly related with the local culture, it is necessary to approach it differently in each setting. In Colombia, as well as in other developing countries, there are few studies on the relationship between alcohol consumption and the risk of chronic diseases⁷. This is due mainly to the lack of valid and high-cost alcohol assessment methods suitable for research in large populations. Moreover, the use of methods with low validity seriously attenuates the associations between nutritional intake and disease in epidemiological studies, a problem known as regression dilution¹¹.

The goal of the present research was to establish the reproducibility and relative validity of two AFQs,

†Correspondence address: Calle 33 Número 29–14, Apartamento 501, Zona postal 3001, AA 260 Bucaramanga, Colombia.

developed to classify subjects according to their alcohol intake, in the adult population of Bucaramanga, Colombia.

Methods

The study was developed in four stages: (1) selection of subjects; (2) application of three semi-quantitative records of alcohol consumption for 30 consecutive days (30-DR); (3) application of two AFQs; and (4) study of the reproducibility and relative validity of the AFQs.

Selection of subjects

Bucaramanga is a medium-sized city, with approximately one million inhabitants. The target population was adults between 20 and 60 years old. We considered subjects as eligible if they admitted to have consumed alcohol in the last month (at least one ounce of any liquor or one beer) and were literate. Through a sampling strategy with multiple stages, 109 alcohol drinkers were selected to represent Bucaramanga by sex and social status¹². Social status was determined by the Municipal Planning Office. This sample, overestimated by 10% to allow for possible losses to follow-up, allowed us to obtain correlation coefficients of 0.40 minimum between the observed consumption and that estimated by the AFQs ($\alpha = 0.05$, $\beta = 0.90$). Every block in the city was numbered and classified according to the six socio-economic strata. From a map of the city, five blocks were sorted out by status (30 in total); a census of eligible subjects was carried out in each household, and one subject was randomly selected per household to complete the sample size for each social stratum. After the selection process, informing the subjects about the objectives of the study and obtaining written consent from the subjects, the survey was applied. The research and the procedures were approved by the Ethics Committee of the School of Health at Universidad Industrial de Santander, Colombia.

Application of three 30-DRs

Each subject completed all three 30-DRs consecutively; the 30-DRs yielded 90 days of records of the drinks consumed and their alcohol content. Before starting the first 30-DR, each subject received training from a nurse or a nutritionist on how to describe the date of consumption, the type of drink (straight or mixed), the volume consumed, the size of glass, the amount of ice used, and how to fill in the daily record for one week. They also received written instructions to be consulted when necessary. Following the training, each subject recorded their alcohol consumption during one week in order to assess their ability to produce complete and clear records. On the first day of follow-up the subjects answered a survey querying sociodemographic data. Expressed like for alcoholic drinks was the self-report of the perception each subject had. During the 30-DR, the nurse and the nutritionist

visited each subject at least once a week. Subjects could consult either the nutritionist or the nurse via telephone at any time of day in order to clarify any doubts.

The content of alcohol in each alcoholic drink, expressed in grams, was obtained from the information provided by the manufacturers (weight–volume). For mixed drinks the content was obtained from the 30-DR according to the contribution of the alcoholic drink in the mix. A composition table with 53 items was derived from the collected information. This table included the name of the beverage or mix, the alcohol content (g) in 100 ml and the total energy, alcoholic and non-alcoholic¹³.

Application of two AFQs

Three months later, after the end of the last 30-DR, two AFQs (AFQ-A; AFQ-B) were applied by interview with a 30-day interval between each one. Each AFQ was applied twice. The application sequence was randomly assigned to each subject to avoid them memorising the responses (bias). The AFQs were given in the following combinations: ABAB, ABBA and BAAB. Both AFQs referred to the previous month's intake.

AFQ-A (53 items)

This questionnaire was developed based on the population's alcohol intake pattern. The check list included the main sources of alcohol, the mixtures made with it and the proportion of the mixer (e.g. cola: 1:1, 1:2 or 2:1). Furthermore, the questionnaire included three intake frequency options (daily, weekly and monthly). Measurement units used for sale or consumption allowed us to establish alcohol intake in a semi-quantitative way.

AFQ-B (five items)

This questionnaire was derived in stepwise strategy using the Max_r method, which aims to maximise the correlation coefficient in the sample while minimising the square error. Max_r achieves this by replacing, in the regression model, the surrogate W_i (the amount of alcohol reached with alcoholic beverages selected by the subject i) by Z_i (the quantity of alcohol in all alcoholic drinks consumed by the subject i). As in classic regression models for each individual the W_i score is used rather than Z_i , a more accurate expression of the between-person variation explained might be R_w^2 ¹⁴. The maximised Pearson coefficients (r) and R_w^2 for each source of alcohol were estimated with Max_r version 2.1¹⁴. Our AFQ-B included the sources of alcohol selected by Max_r in the previous step and was designed following the recommendations of Willett⁷. Questions about usual alcohol intake by source (whether beer or liquor) referred to the previous month (the month before applying the questionnaire). Nine exclusive frequency categories, from 'never' to 'two or more times a day', were used. AFQ-B was derived from data obtained from the semi-quantitative 90 consecutive day record (three 30-DRs) in the same population.

For the three alcohol intake estimation methods (30-DR, AFQ-A, AFQ-B), the individual intake of alcohol per type of source drink was estimated by summing the products of the volume of drink consumed and the quantity of alcohol contained in the beverage, using FoodCalc¹⁵ together with the food composition table¹³. To calculate the alcohol consumed from the AFQ-B, the specific means of alcohol intake were used, by sex, in each of the five items.

Relative validity and reproducibility study

Geometric means and 95% confidence intervals (CI) were used to describe average monthly alcohol intake. The relative agreement between the AFQs and the average alcohol intake from the 90 days of records was assessed using Pearson correlation coefficients, after transformation to the natural log scale. The distribution of alcohol intake was negatively skewed. We also estimated the non-parametric Spearman rank correlation coefficient (r_s) using the original data (before log-transformation). The Pearson and Spearman correlation coefficients measured only the precision of the relationship between the AFQs and 30-DR.

To measure both the precision and the accuracy of the relationship, we used Lin's concordance correlation coefficient (ρ_C)^{16,17}. Lin's coefficient evaluates whether the observed data deviate significantly from the line of perfect concordance (i.e. the line at 45°). If a scatter plot defined by the AFQ and 30-DR values is very tight (precision) and very close to the line of perfect concordance (accuracy), Lin's coefficient will increase towards a maximum value of -1 or $+1$. Correlation coefficients do not provide an appropriate evaluation of relative reproducibility or validity when dietary data are used in categories or groups in the analysis of epidemiological studies¹⁸. The correlation coefficients between the 30-DR and AFQs were adjusted by the energy linked to the alcohol intake. Adjustment for total energy intake was made by using Willett's residual method^{7,19}.

To evaluate the performance of the AFQs in categorising subjects into the same group as the 30-DR we calculated Cohen's weighted kappa statistic (K_w)^{20,21}. Quartile values of the observed alcohol intake distribution of the AFQs and 30-DR were used as cut-off points for cross-tabulation, with weights equal to $1 - |i - j| / (k - 1)$, where i and j index the rows and columns of the table and $k = 4$ (the maximum number of possible ratings).

Correlation coefficients have been criticised as a measure of agreement because they measure the strength of a linear relationship between two variables, but give no information about the magnitude and direction of the difference (absolute bias) between the two variables^{22,23}. To examine the agreement between the AFQs and the 30-DR, we used Bland and Altman's limits of agreement (LOA) method^{22,23}. In addition, we fit a linear regression with the absolute bias as the dependent variable and the mean of AFQs and 30-DR as the independent variable to evaluate if the bias changed significantly with alcohol

intake¹⁷. This analysis was conducted with the log-transformed data and absolute bias and LOA were exponentiated and multiplied by 100 to express the relative increase in intake assessed by the AFQs compared with 30-DR measured intake. Finally, the reproducibility of the 30-DR was evaluated between the first 30-DR and the average of the second and third 30-DRs.

Results

Studied population

A total of 109 subjects (between 17 and 19 per socio-economic stratum) participated in the study, 49.5% (54) were men. The average age was 27.8 (95% CI 25.8, 29.8) years, and was greater for men ($P = 0.06$). The average length of education was 10 years, without any difference by gender ($P = 0.82$). Twelve per cent of the subjects had a previous medical diagnosis (disease or disorder) and 10.0% had followed some sort of diet in the previous year (Table 1). The preference for alcoholic drinks was not different between genders ($P = 0.96$). While the prevailing frequency of consumption per self-report in men was weekly or every two weeks, in women it was monthly. However, the 30-DR showed that, on average, men recorded to have consumed alcohol every 6.7 (95% CI 6.4, 6.9) days, while women recorded alcohol consumption once every 7.3 (95% CI 7.0, 7.7) days ($P = 0.00$).

Follow-up and record of consumption of alcoholic beverages

A total of 167 houses were visited; in 27 of them no eligible subjects were found. With 240 eligible subjects in the remaining 140 houses, a list was made and 109 were selected at random. After training, all subjects were classified as able to record. No subject was lost during the three months of the follow-up, and no record was invalidated due to poor quality. At the end of the three 30-DR, 1324 records were obtained (one per each consumption); about 42% were made on Saturdays, 20% for Fridays and Sundays, and approximately 4% for each of the remaining days. Out of the total of records, 51.1% were made by women.

Alcohol intake and derived energy

The monthly average consumption of alcohol was 226 (95% CI 200, 255) g; for the first month it was 222 (95% CI 181, 262) g, for the second 239 (95% CI 206, 273) g and 398 (95% CI 301, 496) g for the third. Men drank more than women: 312 (95% CI 269, 362) g versus 165 (95% CI 141, 192) g ($P = 0.00$). Based on the number of records and the quantity of alcohol consumed, we established that 61.2% of the population consumed five or more drinks per occasion. Since each subject recorded more than one occasion during the follow-up, we also established that this condition (consumption of five or more drinks of alcoholic beverages and mixers) represented a monthly average energy intake of 8632 (95% CI 7634, 9761) kJ

Table 1 Sociodemographic characteristics

Characteristic	Total (n = 109)	Men (n = 54)	Women (n = 55)	P-value
Age (years)*	27.8 (25.8–29.8)‡	29.8 (26.7–32.8)	25.8 (23.2–28.5)	0.06
Years of education	10.4 (9.7–11.1)	10.3 (9.2–11.4)	10.5 (9.6–11.4)	0.82
Weight (kg)	61.8 (59.6–64.0)	67.8 (64.7–70.9)	55.9 (53.6–58.2)	0.00
Body mass index (kg m ⁻²)	22.6 (21.9–23.2)	23.1 (22.0–24.0)	22.0 (21.1–22.9)	0.10
> 25 kg m ⁻²	25 (22.9)§	14 (25.9)	11 (20.0)	0.31
Actual disorder?	13 (11.9)	4 (7.4)	9 (16.4)	0.15
Some diet?†	11 (10.1)	5 (9.3)	6 (10.9)	0.77
Socio-economic status				0.98
1	19 (17.4)	10 (18.5)	9 (16.4)	
2	19 (17.4)	9 (16.7)	10 (18.2)	
3	19 (17.4)	10 (18.5)	9 (16.4)	
4	19 (17.4)	9 (16.7)	10 (18.2)	
5	17 (15.6)	8 (14.8)	9 (16.4)	
6	16 (14.8)	8 (14.8)	8 (14.4)	

* Mean 27.8 (standard deviation 10.6) years; the distribution of age is asymmetric.

† In the last year.

‡ Mean (95% confidence interval).

§ Frequency (%).

(2062 kcal). Out of this total, 6611 (95% CI 5856, 7465) kJ (1579 kcal) were alcohol-derived. Men consumed more non-alcoholic energy (e.g. derived from cola drinks) than women: 2756 (95% CI 2273, 3330) kJ (658 kcal) versus 1355 (95% CI 1133, 1610) kJ (324 kcal) ($P = 0.00$). Alcohol intake did not differ by socio-economic status for the methods 30-DR ($P > F = 0.78$) and AFQ-A ($P > F = 0.33$), but it did for the AFQ-B ($P > F = 0.00$).

Application of the AFQs

Ninety-six (95% CI 93, 99) days elapsed between the end of the 30-DR and the first application of an AFQ. Each AFQ was applied twice and 60 (95% CI 58, 62) days elapsed between the first and the second application. In the first interview, all of the subjects answered. In the second AFQ-B application 108 answered. To apply the AFQ-A, 25–30 min were needed; between 5 and 10 min for AFQ-B.

AFQ-B design

Of 10 alcoholic drinks consumed by the studied subjects, five were selected by Max_r¹⁴, reaching a maximised r (Pearson) of 0.99 and R^2_{wv} of 99.8.

Reproducibility study

Reproducibility was higher for the AFQs than for the 30-DR. Alcohol intake was greater for all methods when they were applied the second time. The increase was slight for the AFQs (10% and 6%), but was 60% for the 30-DR. The correlation coefficients, which allows one to test reproducibility in variables when they are treated as continuous, can be considered mild and good for the AFQs, whereas for 30-DR they are low. The same occurred when testing reproducibility in variables categorised in quartiles. There was no evidence of bias (LOA) in agreement for the consumption amount in any of the three methods (Table 2). AFQ-A had better relative reproducibility than AFQ-B. The low reproducibility of

the 30-DR was substantially influenced by the third measurement. While the deviation (LOA) between the first and second 30-DR was only 18.4%, between the second and the third it was 42.4%, and between the first and the third it was 60.8%.

Relative validity study

Unlike the previous results, AFQ-B had more relative validity than AFQ-A. The intake of alcohol estimated with the AFQs was greater than that calculated with the 30-DR; 56% for AFQ-A and 11% for AFQ-B. While the difference in alcohol estimated for AFQ-A was positively asymmetric, the average difference estimated for AFQ-B was normally distributed. When evaluating the relative validity of the variables treated as continuous, the correlation coefficients could be considered good for AFQ-B and moderate for AFQ-A. There was no evidence of bias (LOA) in agreement linked to the amount consumed (Table 3) nor an influence of socio-economic status on the relative validity for either of the AFQs. There was an increase when evaluating the agreement degree reached by the AFQs compared with the 30-DR using geometric means: 74% for the AFQ-A, and it was stable for AFQ-B (11%). 'Energy-adjusted' alcohol intake was computed as the residual from the regression model with total energy intake as the independent variable and absolute alcohol intake as the dependent variable. The alcohol residual by definition provides a measure of alcohol intake uncorrelated with total energy intake. Complementarily, the calculation of coefficients based on partition of the components of the variance allows correction for the attenuation originating from the random error (within-person)⁷. Pearson's raw and corrected correlation coefficients for alcohol were good in both methods (0.52 minimum, 0.68 maximum). When adjusted by energy, they were reduced considerably (0.22 minimum, 0.31 maximum). However, in any circumstance they were statistically significant ($P, (r = 0) < 0.001$)

Table 2 Reproducibility study. Correlation coefficients*, between two applications, for three methods of estimating alcohol intake (g)†: a 30-day semi-quantitative daily record (30-DR; reference method) and two alcohol frequency questionnaires (AFQ-A, 53 items; AFQ-B, five items)

Dietary factor	r_{\dagger}	ρ_C §	r_S ¶	K_w	Bias (LOA)**	P -value††
Between 30-DR (reference method)						
Total energy	0.42	0.36	0.41	0.23	-43 (-216, 130)	0.85
Alcohol (g)	0.21	0.14	0.33	0.20	60 (-236, 355)	0.69
Alcoholic energy	0.41	0.35	0.39	0.26	-43 (-214, 128)	0.85
Non-alcoholic energy	0.33	0.30	0.41	0.33	-44 (-288, 200)	0.88
Between AFQ-A (53 items)						
Total energy	0.76	0.75	0.70	0.49	11 (-108, 130)	0.99
Alcohol (g)‡‡	0.75	0.75	0.68	0.50	9.8 (-110, 130)	0.99
Non-alcoholic energy	0.75	0.74	0.73	0.56	15.1 (-114, 144)	0.99
Between AFQ-B (5 items)						
Total energy	0.55	0.53	0.52	0.34	2.1 (-190, 195)	0.97
Alcohol (g)‡‡	0.57	0.55	0.55	0.38	6.1 (-180, 192)	0.96
Non-alcoholic energy	0.13	0.12	0.50	0.33	2.4 (-601, 605)	0.98

* Values log-transformed when necessary.
 † Estimated monthly.
 ‡ Pearson's correlation coefficient. $P_r(r=0) < 0.05$ for all, except for non-alcoholic energy of the AFQ-B.
 § Lin's concordance correlation coefficient.
 ¶ Spearman's correlation coefficient. $P_{r_s}(r_s=0) < 0.05$ for all.
 || Weighted Cohen's kappa for quartiles (weight = $1 - |i - j| / (k - 1)$), where i and j index the row and columns of the two ratings and $k = 4$.
 ** Relative individual intake differences and limits of agreement (LOA) with 95% confidence interval in parentheses (first application minus second application).
 †† Linear trend.
 ‡‡ The coefficients for alcohol are the same as for alcoholic energy.

(Table 3). The agreement between the AFQs for alcohol intake can be considered mild ($K_w = 0.40$) (Table 4).

Discussion

In underdeveloped countries like Colombia, where cardiovascular and chronic diseases have reached numbers considered of epidemic proportions²⁴ and where it is suggested that this cannot be satisfactorily explained with models developed in other societies²⁵, it is

necessary to develop reliable measurement methods suitable for such particular circumstances found in the culture. The development and validation of an AFQ is a difficult task, due in part to the difficulties in obtaining a representative sample of the population in which the AFQ will be applied. Because of its design, our study could only include subjects able to read and write and who were willing and able to complete the alcoholic drink records accurately. Although we used a wide age range, we cannot verify whether the age distribution of our sample was

Table 3 Validation study*. Correlation coefficients† of alcohol intake estimated by two alcohol frequency questionnaires (AFQ-A; 53 items; (AFQ-B; five items) versus the intake calculated from a 30-day semi-quantitative daily record (30-DR; reference method)

Dietary factor	Both applications				First application				Second application			
	ρ_C ‡	r_S §	Bias (LOA)¶	P -value	r_R **	r_A ††	r_{R-A} ‡‡	r_{A-C} §§	r_R **	r_A ††	r_{R-A} ‡‡	r_{A-C} §§
AFQ-A (53 items)												
Total energy	0.44	0.48	-59.2 (-191, 72)	0.73	0.54		0.57	0.57	0.59		0.63	0.63
Alcohol (g)¶¶	0.46	0.50	-56.4 (-185, 72)	0.75	0.56	0.29	0.58	0.31	0.60	0.25	0.63	0.27
Non-alcoholic energy	0.38	0.41	-69.0 (-228, 90)	0.73	0.46	0.25	0.49	0.26	0.51	0.22	0.54	0.23
AFQ-B (five items)												
Total energy	0.61	0.62	-9.9 (-144, 124)	0.95	0.63		0.71	0.71	0.52		0.59	0.59
Alcohol (g)¶¶	0.59	0.60	-11.0 (-146, 124)	0.94	0.60	0.20	0.68	0.22	0.52	0.23	0.58	0.26
Non-alcoholic energy	0.33	0.63	33.7 (-268, 335)	0.71	0.45	0.13	0.81	0.23	0.28	0.08	0.51	0.15

* Estimated monthly alcohol intake (g).
 † Values log-transformed when necessary.
 ‡ Lin's concordance correlation coefficient.
 § Spearman's correlation coefficient. $P_{r_s}(r_s=0) < 0.05$ for all.
 ¶ Relative individual intake differences and limits of agreement (LOA) with 95% confidence interval in parentheses (30-DR minus AFQ).
 || Linear trend.
 ** Raw Pearson correlation coefficient. $P_r(r=0) < 0.05$ for all.
 †† 'Energy-adjusted' Pearson correlation coefficient. $P_r(r=0) < 0.05$ for all, except for non-alcoholic energy of the AFQ-B.
 ‡‡ Raw and corrected Pearson correlation coefficient; $r_1 = r_o \sqrt{1 + (\lambda_x / n_x)}$, where r_1 is the true correlation, r_o is the observed correlation, λ_x is the ratio of the within- and between-person variance for x and n_x is the number of replicates per person for the variable x
 §§ 'Energy-adjusted' and corrected Pearson correlation coefficient.
 ¶¶ The coefficients for alcohol are the same as for alcoholic energy.

Table 4 Validation study*. Ratio† of alcohol intake estimated by two alcohol frequency questionnaires (AFQ-A, 53 items; AFQ-B, five items) to that calculated from a 30-day semi-quantitative daily record (30-DR; reference method), and classification of subjects by risk according to the AFQs

Dietary factor	AFQ-A (53 items)		AFQ-B (five items)	
	AFQ-A/30-DR (%)	K_w ‡	AFQ-B/30-DR (%)	K_w ‡
Total energy	179 (174, 185)	0.32	110 (106, 115)	0.38
Alcohol (g)§	174 (169, 180)	0.40	111 (106, 116)	0.40
Non-alcoholic energy	198 (194, 202)	0.29	71 (59, 86)	0.43

* Estimated monthly alcohol intake (g).

† Based on geometric means, with 95% confidence interval in parentheses.

‡ Weighted Cohen's kappa for quartiles (weight = $1 - |i - j| / (k - 1)$, where i and j index the row and columns of the two ratings and $k = 4$).

§ The coefficients for alcohol are the same as for alcoholic energy.

similar to that of the general population owing to a lack of updated demographic data. The age average was probably younger and the level of education superior in our sample. Therefore, the validity of the AFQs may be lower if administered to subjects older than 40 years of age or with a lower level of education, even though the fraction of subjects without primary school (5.5%) was close to that in the target population (4.8%)²⁶.

The sampling strategy, together with the 30-DR (considered the reference method) and the quality of those records, allow us to suppose that the data with which the AFQs were compared (30-DR) captured the within-person and between-person variation²⁷. However, given the low reproducibility achieved by the 30-DR, it is possible that we may have to record alcoholic drinks intake in non-sequential periods longer than a month (30-DR), or carry out the 30-DR for more than three months, or finally to question the usefulness of the 30-DR as a reference method. In any case, to carry out reproducibility research is difficult, even more so when there is high variation in alcohol intake over a fixed period of time (month, week, weekend, etc.). This also depends on the amount and quality of the subject's help. This, together with what was discussed previously, can be considered a limitation of the research.

It is well known that the dietary habits or the alcohol intake recorded could be underestimated or overestimated due to fatigue or modification in intake habits⁷. There is evidence of this, since the average consumption for the third 30-DR was higher than in the two previous ones.

Traditionally, a reproducibility and validation study should have one measure before and another one after applying the reference method²⁸. Nevertheless, it is reasonable to think that for alcohol intake research particularly, which has social and moral implications in practice, it is only possible to correctly measure this exposure after previous sensitive interventions, which could yield artificial accuracy. On the other hand, it is also possible to reasonably argue that, after three months of finishing the 30-DR, the memory might be blurred and

therefore artificially low correlations between the AFQs and 30-DR could be obtained. In any case, our results are half way between a positive and a conservative position on the real correlation between the AFQs and 30-DR. Moreover, the order of application of the AFQs (ABAB, ABBA and BAAB), their random assignment and the time between applications allowed us to control for a possible bias of memory.

Since the number of drinks per occasion is not constant when a subject consumes alcoholic beverages (only 8%), our findings allow us to state that, in this population, methods widely used to classify subjects as being at risk according to the number of drinks are inefficient^{29,30}. For this reason, our results are especially relevant.

In nutritional epidemiology, it is desirable to have replicated methods, even in circumstances where variation of the factor under study is high^{7,28,31,32}. Our results on reproducibility show it beyond any doubt; while the Pearson correlation coefficient for alcohol was 0.21 for the 30-DR, it was between 0.57 and 0.75 for the AFQs. We observed the same for the total energy, alcoholic energy and non-alcoholic energy (Table 2).

The relative validity of the AFQs against the 30-DR was established in several ways. AFQ-B had better performance than AFQ-A, yielding a more reliable coefficient when the variables were handled as continuous (ρ_c : 0.59 vs. 0.46 for alcohol). It also had greater precision (LOA). The poor performance of AFQ-B when reproducing non-alcoholic energy calculated by the 30-DR was predictable and is due to the check list characteristics (five items), which did not include mixed drinks (Table 3). The greatest difference observed, between the AFQ-A and the 30-DR (56%), is most likely due to the greater number of items (53): 'the more you ask, the more overreporting you get'¹⁰. When evaluating the capacity to correctly classify subjects according to intake in quartiles, the two AFQs performed the same ($K_w = 0.40$) (Table 4). Since the difference between the AFQ-B and the 30-DR is only 11%, it could be a useful tool for determining the consumption of alcohol (level of intake) in cross-sectional surveys or population interventions.

The stability of the results related to repeat application of a test is also an indirect reproducibility measure. The second application of both AFQs did not substantially change the estimated values for total energy, non-alcoholic energy or alcohol (g), which represents an additional advantage of these AFQs versus the 30-DR. Pearson's raw and attenuation-corrected correlation coefficients are good, comparing those estimated for these three variables versus those calculated by the 30-DR.

We calculated the correlation adjusted for energy, and the coefficients dropped substantially (Table 3). This is mainly due to the high correlation between alcohol and total energy, since 77% comes from the alcohol. Despite this, values of the energy-adjusted coefficients are within the range found in other studies that included food

consumption of the diet^{10,33,34}. It is possible that alcohol consumption patterns like the Mediterranean, characterised by moderate daily intake and related to meals, reproduce the correlations between alcohol and energy observed in this research. Traditionally the correlation between total diet energy and alcohol intake is low³⁵. We presuppose that it is not necessary to adjust for energy intake in studies of alcohol intake only and that, in validation of such studies, raw correlation coefficients can express agreement between the test and the reference method. The raw or simply corrected correlation coefficient may be the best expression of validity when the alcohol consumption pattern is characterised by a high weekly consumption, not associated to foods, like the one reported in this study.

In conclusion, we developed a study to evaluate the reproducibility and validity of two AFQs in a population where alcohol intake variation is high, due to the consumption pattern or habit. Since the relative validity of AFQ-B is very acceptable and this questionnaire has only five items, it becomes the best option. AFQ-B is applied quickly and is highly cost-effective. Our study shows that these AFQs may be useful as an alcohol assessment tool for future epidemiological studies in the population of Bucaramanga. However, careful consideration should be given to issues related to study design in order to control for the effect of non-differential measurement error on alcohol intake. These AFQs could also be used as a starting point for the development and validation of AFQs in other populations in Colombia, and to monitor changes across populations and through time.

Acknowledgements

Funding: This work was financed by the Colombian Institute for the Development of Science and Technology 'Francisco José de Caldas' (COLCIENCIAS; grant no. 127–2002, code 1102-04-11 720) and Universidad Industrial de Santander (UIS).

Author contributions: Both authors participated in the conceptualisation and design of the study. O.F.H. supervised the development of the study, conducted most of the data analysis and wrote the manuscript. M.F.A. was in charge of the fieldwork, designed the questionnaires, participated in the data analysis and revised the manuscript.

References

- 1 Howson CP, Reddy KS, Ryan TJ, Bale JR, eds. *Control of Cardiovascular Diseases in Developing Countries. Research, Development and Institutional Strengthening*. Washington, DC: National Academy Press, 1998.
- 2 Omran AR. The epidemiological transition: a theory of the epidemiology of population change. *Milbank Fund Quarterly* 1971; **49**: 509–38.
- 3 Murray CLJ, Lopez AD. Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. *Lancet* 1997; **349**: 1498–504.
- 4 de Torres GY. Alcohol: prevalencia de consumo y dependencia en Colombia. *Revista de Medicina CES, Medellín* 1999; **12**: 1–9.
- 5 Torres Y, Posada V, Rojas MC. *Estudio Nacional de Salud Mental y Consumo de Austancias Sicoactivas*. Bogota: Ministerio de Salud, 1993.
- 6 Munera R, Vélez N, Zuluaga MT. Prevalencia del consumo del alcohol en la población escolarizada del municipio de Sabaneta, 1992–1993. *Revista de Medicina CES, Medellín* 1994; **8**: 73–80.
- 7 Willett WC, ed. Food-frequency methods. In: *Nutritional Epidemiology*. New York: Oxford University Press, 1998; 74–100.
- 8 Navarro A, Osella AR, Guerra V, Munoz SE, Lantieri MJ, Eynard AR. Reproducibility and validity of a food-frequency questionnaire in assessing dietary intakes and food habits in epidemiological cancer studies in Argentina. *Journal of Experimental & Clinical Cancer Research* 2001; **20**: 365–70.
- 9 Slater B, Philippi ST, Fisberg RM, Latorre MRDO. Validation of a semi-quantitative adolescent food frequency questionnaire applied at a public school in Sao Paulo, Brazil. *European Journal of Clinical Nutrition* 2003; **57**: 629–35.
- 10 Feunekes GI, van't Veer P, van Staveren WA, Kok FJ. Alcohol intake assessment: the sober facts. *American Journal of Epidemiology* 1999; **150**: 105–12.
- 11 Day N, McKeown N, Wong M, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *International Journal of Epidemiology* 2001; **30**: 309–17.
- 12 Departamento Administrativo Nacional de Estadística (DANE). *Censo de Población Colombia*. Bogota: DANE, 1993.
- 13 Herrán OF, Bautista LE, Quintero DC. *Tabla de Composición de Alimentos Consumidos en Bucaramanga*, 2nd ed. Bucaramanga, Colombia: Centro de Investigaciones Epidemiológicas, Universidad Industrial de Santander, 2003; 69–72.
- 14 Mark SD, Thomas DG, Decarli A. Measurement of exposure to nutrients: an approach to the selection of informative foods. *American Journal of Epidemiology* 1996; **143**: 514–21.
- 15 Lauritsen J. *FoodCalc 1.3; Diet, Cancer and Health project at the Danish Cancer Society*. Copenhagen: Danish Cancer Society, 1998.
- 16 Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; **45**: 255–68.
- 17 Lin LI-K. A note on the concordance correlation coefficient. *Biometrics* 2000; **56**: 324–5.
- 18 Masson LF, McNeill G, Tomany JO, Simpson JA, Peace HS, Wei L, *et al*. Statistical approaches for assessing the relative validity of a food-frequency questionnaire: use of correlation coefficients and the kappa statistic. *Public Health Nutrition* 2003; **6**: 313–21.
- 19 Rosner B, Willett WC. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *American Journal of Epidemiology* 1988; **127**: 377–86.
- 20 Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960; **20**: 37–46.
- 21 Cohen J. Weighted kappa: nominal scale agreement with a provision for scaled disagreement or partial credit. *Psychological Bulletin* 1968; **70**: 213–20.
- 22 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307–10.

- 23 Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 1999; **8**: 135–60.
- 24 Ministerio de Salud. *La Carga de la Enfermedad en Colombia*. Bogota: Ministerio de Salud, 1994; 65–73.
- 25 Peña M, Ballacao J. *La Obesidad en la Pobreza; Un Nuevo Reto para la Salud Pública*. Publicación Científica No. 576. Washington, DC: Organización Panamericana de la Salud, 2000.
- 26 PROFAMILIA. *Encuesta Nacional de Demografía y Salud. Resumen Región Oriental, 2000*. Bogota: PROFAMILIA, 2000.
- 27 Marquis G. Método de pesos y medidas. En: *Manual de Encuestas de Dieta*. Perspectivas en Salud Pública No. 23. México: Instituto Nacional de Salud Pública, 1996; 147–71.
- 28 Cade J, Thompson R, Burley B, Warm D. Development, validation and utilisation of food-frequency questionnaires – a review. *Public Health Nutrition* 2001; **5**: 567–87.
- 29 Wechsler H, Isaac N. 'Binge' drinkers at Massachusetts colleges; prevalence, drinking style, time trends, and associated problems. *Journal of the American Medical Association* 1992; **267**: 2929–31.
- 30 Wechsler H, Dowdall GW, Davenport A, Castillo S. Correlates of college student binge drinking. *American Journal of Public Health* 1995; **85**: 921–6.
- 31 Margetts BM, Nelson M. *Design Concepts in Nutritional Epidemiology*, 2nd ed. London: Oxford University Press, 1996; 123–69.
- 32 Kraemer HC. *Evaluating Medical Tests*. London: Sage Publications, 1992; 63–95.
- 33 Decarli A, Franceschi S, Ferraroni M, Gnagnarella P, Parpinel MT, La Vecchia C, *et al.* Validation of a food-frequency questionnaire to assess dietary intakes in cancer studies in Italy: results for specific nutrients. *Annals of Epidemiology* 1996; **6**: 110–8.
- 34 Ferraroni M, Decarli A, Franceschi S, La Vecchia C, Enard L, Negri E, *et al.* Validity and reproducibility of alcohol consumption in Italy. *International Journal of Epidemiology* 1996; **25**: 775–82.
- 35 Johansson I, Hallmans G, Wikman A, Biessy C, Riboli E, Kaaks R. Validation and calibration of food-frequency questionnaire measurements in the Northern Sweden Health and Disease cohort. *Public Health Nutrition* 2002; **5**: 487–96.