Journal of Official Statistics
Vol. 8. No. 3, 1992, pp. 327–350
© Statistics Sweden

Self-Reported Consumption Measures in Sample Surveys: A Simulation Study of Alcohol Consumption

John C. Duffy[1] *and Timo Alanko*[2]

Abstract: Four methods of measuring alcohol consumption by self-report in sample surveys are compared on the basis of simulation studies of the drinking process in human populations. Properties of population estimates and the classification of individuals are considered and the implications of the results for population surveys and epidemiological studies discussed.

Key words: Self-report; measurement error; alcohol surveys; renewal processes; simulation.

1. Introduction

Many sample surveys include questions attempting to elicit information about the consumption of alcoholic beverages by respondents. In the United Kingdom for example, Government-sponsored population surveys of national drinking habits have been carried out several times in the last decade (Breeze 1985; Goddard 1991; Goddard and Ikin 1989; Wilson 1980), while in Finland the Drinking Habits Survey is conducted every eight years (Simpura 1987). Other government surveys, for example, the Family Expenditure Survey (OPCS 1990a) and the General Household Survey (OPCS 1990b) in the UK and similar Scandinavian surveys (for instance, CSOF 1985) also commonly include questions about alcoholic beverage consumption and purchases. Similar surveys are conducted in most developed countries, and in many other parts of the world.

Population surveys, albeit usually on a smaller scale, are also frequently used by researchers in the field of social and medical problems related to alcohol, and in studies of the relationship between alcohol and disease risk in clinical epidemiology; the self-reported alcohol consumption of patients is often compared with that of controls.

The early finding that population estimates from sample surveys did not begin to approach known population consumption (from taxation and sales data, for example) stimulated interest in sources of bias in

[1] Senior Lecturer, Department of Mathematics and Statistics, University of Edinburgh, on secondment as Director, Statistics and Information, Alcohol Research Group, Department of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, Scotland.
[2] Senior Research Officer, Computing Service, Technical Research Centre of Finland, Otakaari 11, 02150 Espoo, Finland.

Acknowledgements: The authors are grateful to the Finnish Foundation for Alcohol Studies, the Social Research Institute for Alcohol Studies, Helsinki, and the Portman Group, London, for assistance with research travel and the provision of research facilities.

survey measurement. These may be considered under the headings of sample frame defects, response bias, which would include interviewer effects and deliberate under-reporting, memory bias (forgetting), and measurement bias, which forms the subject of the present work. Alternative methods of measurement and administration of questionaires have been studied, and some insights gained into the relative contributions of these sources of bias as causes of under-coverage (Alanko 1984; Duffy and Waterton 1984; Waterton and Duffy 1984; Wilson 1981; Midanik 1982; Poikolainen and Kärkkäinen 1983, 1985).

At the same time various forms of the distribution of consumption theory (Ledermann 1956; Duffy 1986) stimulated interest in more theoretical aspects of the alcohol consumption of individuals, in particular stochastic modelling of the consumption process in terms of renewal theory (Ekholm 1968; Alanko 1984). This approach considers the consumption behaviour of individuals as a process of events occurring in time, the occasions of consumption, with each of which is associated an amount consumed. The total consumption of a respondent over a particular time period may thus be modelled as a random sum.

Computer simulation of this renewal theory model offers a means of addressing some basic problems of measurement in sample surveys. The strategy adopted in the present work involved defining the methods of measurement to be considered, simulating the appropriate drinking processes and examining the properties of the resulting estimates.

The measurement methods investigated here fall into two natural categories. The first category, of 'retrospective diary' methods, attempts to elicit information about the actual drinking behaviour of individuals over a period of time. Respondents are questioned about the beverages and amounts they drank on each recalled occasion of drinking. The second category, which may be subject to interpretation by respondents, is the 'summarial' approach in which respondents are asked to report their usual drinking habits in terms of frequency of occasions of consumption and amounts consumed per occasion. Both of these types of method are in widespread use in Government and other surveys related to alcohol consumption (see Section 4).

Temporal variation in drinking behaviour by respondents is the principal source of concern regarding measurement methods. In the simulation study two population drinking processes were modelled, one of beer drinking by males based on parameter estimates from a Finnish study (Simpura 1987), and the other of all male drinking based on estimated parameters from a survey conducted in England and Wales (Wilson 1980). The populations corresponding to these surveys have a low frequency and high frequency, respectively, of drinking occasions over time.

2. Aims of Measurement

2.1. *Mean consumption*

There are at least three possible aims of measurement in surveys of alcohol consumption. We may treat first the estimation of population mean consumption. In the case of an entire country, which levies taxes on alcoholic drinks and has a reasonably comprehensive statistical reporting system on taxation, this objective is of little interest. However domain estimation of mean consumption for subgroups of the populations may be of considerably more value and it is clear that considerations of measurement bias and precision in estimating mean consumption will be relevant to almost all applications of survey methods to alcohol consumption behaviour.

2.2. *Exceedance proportions*

The estimation of exceedance proportions is also relevant, in connection with estimating the prevalence of drinking at levels which might result in adverse health consequences. There are three aspects which could be distinguished here, one being the accuracy of estimation of exceedance proportions, and the second being the usefulness of the survey instrument as a screening tool. In simple terms the questions are 'does the method give a good estimate of the proportion of the population consuming in excess of a particular quantity of alcohol per year?' and 'are those individuals identified by the survey as consuming more than the threshold amount in the particular period of interview actually consuming at this level over the longer period?' It is of course possible for the first question to be answered affirmatively without implying the second.

The sensitivity of a screening instrument is usually defined as the proportion of positive cases correctly identified as such, while the specificity is the proportion of identified cases who actually are positive (the proportion of 'true positives' among all positives). In terms of alcohol consumption, the sensitivity of a measurement method at a given threshold of consumption may be estimated by the proportion of individuals who actually exceed the threshold who are classified as such by the measure. The specificity is estimated as the proportion of those classified as exceeding the threshold who in fact do exceed the threshold. Both of these quantities are therefore relevant to the comparison of measures with respect to their usefulness in screening.

The third topic in this connection is the examination of the distribution of consumption. Theories relating average consumption in a population to levels of harm usually rely on the distribution of consumption being of a particular form, and in practice, when empirical data are used to address this question, the usual approach is to consider the distribution of amounts consumed in a period among only those respondents who consumed some alcohol in the period. Thus the simulation also examines the extent to which the results of such procedures resemble the actual distribution of consumption over a longer period.

2.3. *Relationship with other factors*

A further important application of survey methods in this field is the investigation of the association between consumption of alcohol and experience of health, social and other problems. In some instances the survey may contain questions relating to possible alcohol-related consequences, while other applications such as case-control or prospective epidemiological studies may rely on external classifications of the problem. The questions of interest in this connection relate to the ordering properties of the measurement methods applied to individuals as well as to their accuracy on a population basis. A method of measurement which results in a biased estimate of the total consumption parameter of each individual in the sample will, if the resulting measurement is monotonically related to the true measurement with no error variation, preserve the rank order of individuals in terms of their consumption. The extent of error variation in the relationship between actual and measured values will determine the extent to which the measured rank order differs from the true ordering.

As the methods are less reliable in ordering the individuals so the association between consumption and the other variables is attenuated or obscured, while if the ordering is satisfactory but the measurement is biased, the estimated parameters of the relationship will be incorrect. For measurements biased downwards the relationship between consumption and problems will appear steeper than it really is, while for measurement methods biased upwards the reverse phenomenon will occur.

3. Individual Drinking Processes and Their Population Distributions

A description of population drinking processes, and aspects of the model to be used in the simulation will be a useful preliminary to discussion of the methods of measurement in common use, considered in the next section.

It should be clear from the discussion in Section 2 that estimation of the consumption of an individual or a population in the time period to which questioning is related (the reference period) is not the primary aim of most alcohol surveys. In some cases the results from a survey of consumption over a short period are used to estimate consumption over a longer period (the estimation period) of a population or domain (usually by the simple expedient of multiplying the relevant mean response by the ratio of the lengths of the two periods). In other cases, the sample proportion of individuals consuming more than a fixed amount during the reference period is used to estimate the average population proportion consuming in excess of this rate over a much longer period. Thus the results of sample surveys in terms of drinking behaviour over a particular reference period are used to estimate more permanent or temporally persistent attributes of the behaviour in the population.

Further, it may be noted that non-stationarity (either temporal trends or cyclical variation) in population consumption behaviour would necessitate the use of time as a dimension of the sampling scheme, but this is rarely implemented in practice (see, however, Lemmens and Knibbe 1991). For this reason, a stationary model of drinking behaviour was adopted for simulation purposes, which will therefore favour survey methods involving the use of short reference periods for longer estimation periods.

The drinking behaviour of an individual respondent over time is assumed to be governed by parameters corresponding to his or her mean frequency of occasions of drinking, the mean amount consumed per occasion and the variability of the amount consumed per occasion. The intervals between drinking occasions for an individual i were taken to follow the exponential distribution with a stable long-term parameter λ_i unique to the individual concerned. Accordingly, the number of drinking occasions within a time period of length t undertaken by individual i follows the Poisson distribution with parameter $\lambda_i t$.

The amounts consumed by individual i on each occasion were modelled by the inverse Gaussian distribution, with mean μ_i independent of λ_i, and the amounts consumed on different occasions of drinking by individual i were assumed independent. The dispersion parameter of the within-individual distribution of amounts per occasion was set on the basis of a secondary analysis of empirical data from the 1984 Finnish drinking survey (Simpura 1987), which showed that the coefficient of variation of this process was roughly constant.

To model the differences between individuals, the individual level frequency (rate) parameter was assumed to follow an inverse Gaussian distribution in the Finnish model and a gamma distribution in the case of England and Wales. The difference in the distributional form of these mixing distributions between the countries was based on examination of survey drinking frequency distributions from respective countries. This leads to compound drinking frequency models (negative binomial and Poisson-inverse Gaussian), well-known from buying behaviour literature (e.g., Sichel 1982, Goodhardt, Ehrenberg, and Chatfield 1984 and references therein). Inverse Gaussian distributions were used for both countries to express variability between mean amounts consumed. The parameters of the mixing distributions reported in this study were chosen by the method of moments to correspond to survey distributions from respective countries (Wilson 1980; Simpura 1987).

For details of the mathematical forms of the distributions employed see e.g., Johnson and Kotz (1970a; 1970b).

4. Methods of Measurement

4.1. Last week's consumption

The first three methods to be described may all be considered as 'retrospective diary' approaches, in which respondents are asked to detail their consumption of alcohol in the immediate past. As the name implies, in the 'last week's consumption' method respondents answer questions about each occasion of alcohol consumption during the seven days immediately prior to the interview, usually in reverse temporal order. Consumption is recorded by beverage type, and amount of each beverage consumed on each occasion. Interviewers are often advised to provide memory prompts to respondents who have difficulty recalling their activities on any particular day.

Questioning in reverse temporal order is thought to reduce memory problems, and the concrete nature of the information required allows less scope for question wording effects and idiosyncratic interpretation by respondents. Last week's consumption has been the method of choice in many UK government surveys, as well as surveys more specifically in the field of alcohol research. Wilson (1981), in a discussion of methodological aspects of the method, also provides a detailed description of its practical implementation in government surveys.

4.2. Survey period method

Once again the occasions of consumption for each respondent are investigated, but in this case after the respondent has answered a question about his usual frequency of drinking. The frequency category of the respondent is used to determine the length of the survey period to be investigated, in accordance with Table 1 (based on the work of Mäkelä (1978) as reported by Simpura (1987), and extended to a most frequent category of more than once per day). The minimum period of investigation is one week, but in appropriate cases this is extended to longer periods, chosen to contain roughly four occasions of consumption. One reason for using this method is to avoid the recording of no consumption for individuals who had consumed no alcohol in the seven days before interview. Consumption data are obtained even for infrequent drinkers, in contrast to the last week's consumption method.

A difficulty which will be discussed later in connection with summarial measurement relates to the way in which respondents interpret the question regarding usual frequency of drinking in terms of the parameters underlying their actual drinking behaviour, and how they then categorise themselves.

Table 1. Drinking frequencies, frequency intervals, survey period lengths and corresponding frequency multipliers for Q-F measurement

Nominal frequency	Interval (occasions/yr)	Survey period length	Frequency multiplier for daily total
less than once annually	0	12 months	1/365
once or twice annually	1–2	12 months	1.5/365
3–4 times annually	3–4	12 months	3/365
roughly once in 2 months	5–8	8 months	6/365
roughly once monthly	9–17	4 months	12/365
twice monthly	18–37	2 months	24/365
once weekly	38–77	4 weeks	52/365
twice weekly	78–155	2 weeks	130/365
4–5 times weekly	156–285	1 week	240/365
daily	286–456	1 week	1
more than daily	457+	1 week	547/365

4.3. *Twenty-four hour recall*

In this method respondents are asked about their consumption on the day preceding interview, which is taken to include the early hours of the morning of the day of interview. Stats MR, a leading British market research company providing services to the drinks industry, has been developing this technique as a new survey product, and similar methods have been used in epidemiological studies related to alcohol and illness (Gordon et al 1981). The putative advantages of this method are that memory problems are likely to be less than with methods requiring recall of a longer time period, and that the cumulative concealment which might occur with retrospective questioning as respondents realise they may be reporting socially undesirable large aggregate amounts over the entire retrospective time period is avoided.

4.4. *Quantity – frequency (Q-F) measurement*

There are a variety of measures which attempt to elicit summary information concerning respondents' drinking habits. These methods have been widely used, modified and elaborated in the United States, and a comprehensive summary of measurements in this tradition has been provided by Room (1985). Although considerably more sophisticated and complex summarial methods have been employed in U.S. surveys, the present work examines the more straightforward approach implemented in the Scandinavian Drinking Survey (Simpura 1981).

Respondents are asked to classify their drinking habits in terms of usual frequency (Table 1) and quantity of each of the three main types of alcoholic beverages as shown in Tables 2 and 3, based on Simpura (1981). The total annual consumption of each beverage by respondents may be estimated by multiplying appropriate coefficients corresponding to each category, and total consumption obtained as the sum over beverage types.

Since summarial methods do not explicitly require recall of actual occasions, it may be inferred that forgetting is likely to be less of a problem than with retrospective diary methods. However, there are other complications. Response category effects in Q-F measurement are well-known, and discussed by Sudman and Bradburn (1982) and Poikolainen and Kärkkäinen (1985).

Table 2. Drinking quantities, quantity intervals and corresponding quantity multipliers for Q-F in simulation of Finnish beer consumption

Nominal consumption (cl of beer)	Interval cl of alcohol	Quantity multiplier for daily total (cls alc)
Less than 1/2 bottle (20 cl)	0–1.47	0.900
About 1/2 bottle (33 cl)	1.48–2.024	1.485
1–2 1/2 bottles (50 cl)	2.025–2.474	2.250
2 1/2 bottles (66 cl)	2.475–3.374	2.970
3 1/2 bottles (100 cl)	3.375–5.624	4.500
4–5 1/2 bottles (150 cl)	5.625–7.874	6.750
6–9 1/2 bottles (250 cl)	7.875–14.174	11.250
10+ 1/2 bottles (350 cl)	14.175+	15.75

Equally important, if not more so, is the problem of how individuals categorise themselves. A respondent would be extremely unlikely to conceive of his or her drinking as a stochastic process, and to report mean occasion or amount parameters. On the other hand individuals could be expected to perform some kind of integration of their behaviour over a period of time to decide which category applied to them. Among the more credible possibilities are that individuals would classify themselves in accordance with the modal value of their occasion and consumption distribution, but this seems rather artificial; alternatively respondents could seek to remember their behaviour over some interval of time and answer on the basis of the mean, mode or some other aspect of their actual experience. The length of this interval would of course have implications for the stability of categorisation, and there is at present no way of estimating what this might be.

Table 3. Quantity intervals and multipliers for Q-F measurement in simulation of consumption by England and Wales men

Quantity upper bounds (units)	Quantity multiplier (units)
< 1.5	1
< 2.5	2
< 4.0	3
< 6.0	5
< 8.5	7
< 12.5	10
< 20	15
20+	25

5. Simulation of Drinking Records and Methods of Measurement

In order to examine the properties of the types of measurements described above, computer simulation of drinking records of samples of hypothetical individuals, drawn from populations similar to Finnish and England and Wales drinking populations, was programmed.

5.1. *Simulation flow*

The steps used in simulating the drinking records and the determination of the measures on the basis of the drinking record and the parameters can be described schematically as follows:

Step 0. Input the mixing distribution parameters, sample size (number of individuals to be simulated), number of Monte Carlo drinking histories to be simulated per individual, etc.

Step 1. Generate for the ith individual a drinking frequency parameter λ_i and a mean amount parameter μ_i from the mixing distributions.

Step 1.1. Determine and save F of Q-F from the drinking frequency parameter. Determine and save Q of Q-F from the mean amount parameter. Determine the length of the reference period for the survey period method.

Step 2. Start generating *j*th drinking history for individual *i*.

Step 2.1. Generate distance (in reverse time) from the previous occasion from the exponential distribution (using the drinking rate parameter). Generate amount consumed on the occasion from the inverse Gaussian distribution (using μ_i as amount parameter and a fixed coefficient of variation).

Step 3. Test if the occasion generated is:

a. Outside 24 h limit (24 h recall).

b. Outside one week limit (last week's consumption).

c. Outside survey period limit determined in 1.1 (survey period method).

d. Outside 8 week limit (Q-F averaging method).

Step 3.1. If answer to any one of 3a–3d is yes, for the first time, determine and save the measure in question from previously recorded (step 4) occasions. Continue to step 4.

Step 3.2. If answer to all of 3a – 3d is yes, output results for the *j*th Monte Carlo run of individual *i*. Return to Step 2. If last Monte Carlo run, return to step 1 for the next individual. If last individual, stop simulation (step 5).

Step 4. Record (accumulate) the occasion and amount generated. Return to step 2.1 for the next occasion.

Step 5. End simulation.

5.2. *Simulation details*

Step 0. Samples of 100, 200, and 1000 individuals were generated both for the Finland and England and Wales simulation models. For each sample size and each individual, 100 Monte Carlo drinking histories were generated. The sample sizes used reflect the range of typical values for survey estimation in total populations and domains, while 100 Monte Carlo runs provides reasonable estimates of summary statistics of interest, including variability.

Finnish beer drinking simulation was based on parameter values estimated directly from the 1984 Finnish Drinking Habits Survey data, using the Poisson-inverse Gaussian model. The mean weekly drinking frequency in the male population was 0.7, with standard deviation 1.2, and the mean amount per occasion 4.0 cl ethanol with standard deviation 4.1. The within-individual coefficient of variation of the amounts was roughly constant and equal to 0.8.

The England and Wales simulation was based on measurement of consumption in terms of units of alcohol, one unit being approximately equal to one half-pint of beer, one standard measure of spirits and one standard glass of wine. Parameter values were set to correspond to the results of the survey reported by Wilson (1980). The mean weekly drinking frequency in the population was taken as 4.8 with standard deviation 0.6, which determines the parameters of the gamma distribution as $\alpha = 2.82$, and $\beta = 1.7$. The mean amount per occasion was 4.1 units and the standard deviation of mean amounts per occasion was based on the assumption of a constant within-individual coefficient of variation of 0.4.

Two approaches to Q-F measurement were implemented, following the ideas of Section 4. The first method involved expressing the frequency and quantity categories as intervals on the appropriate scales of measurement, as in Tables 1, 2, and 3. For each simulated individual the within-individual distributions of occasions and amounts were applied to calculate the probabilities associated with each interval. The individuals were then classified in accordance with the largest value of these probabilities.

The second approach was to simulate an eight week drinking process for each individual, and on the basis of the total number of occasions in the eight weeks the value of frequency was calculated in accordance with Table 4. The average amount per occasion over the eight weeks was used to assign usual quantity on the basis of the intervals in Table 2 in the simulation of Finnish beer drinking and Table 3 in the England and Wales consumption simulation. Different results would be obtained with different length of the period over which consumption is integrated, and eight weeks was chosen simply to assess whether there were any practical differences between the two approaches.

Step 1. Amount and frequency parameters from the appropriate inverse Gaussian distribution using the method of Michael, Schucany, and Haas (1976) as given in Devroy (1986) were generated. In the case of the England and Wales simulation, the frequency parameter was generated using the gamma distribution (Fishman 1978).

Step 1.1. To determine Q-F measures, the first method involved expressing the frequency and quantity categories as intervals on the appropriate scales of measurement, as in Tables 1, 2 and 3. For each simulated individual the within-individual distributions of occasions and amounts were applied to calculate the probabilities associated with each interval. The individuals were then classified in accordance with the largest value of these probabilities. For the survey period method, the generated frequency parameter was used to classify the simulated individual and to determine the length of the reference period according to Table 1.

Table 4. Frequency intervals and multipliers for eight-week averaging process

Number of occasions in 8 week period	Frequency multiplier (daily consumption)
0	3/365
1	6/365
2–3	12/365
4–5	24/365
6–11	52/365
12–23	130/365
24–47	240/365
48–71	1
72+	547/365

Step 2.1. The inverse transformation method was used to generate exponential observations. For the inverse Gaussian see step 1.

Step 3.1. For last week's drinking from step 3a the weekly amount was obtained by summing the amounts. For the 24-hour recall and survey period measures from 3a and 3c, the amounts cumulated were multiplied by an appropriate constant to obtain a measurement-specific estimate of the average weekly total of the individual. The second, averaging, Q-F approach from step 3.1.d, used the total number of occasions in the eight weeks and the value of frequency was calculated in accordance with Table 4. The average amount per occasion over the eight weeks was used to assign usual quantity on the basis of the intervals in Table 2 in the simulation of Finnish beer drinking and Table 3 in the England and Wales consumption simulation. The parameters of each simulated individual from step 1 were used to calculate a true average weekly consumption of the individual for comparative purposes.

5.3. *Analysis and presentation of measures obtained*

Several aspects were examined for each method. Estimates of sample means and between-sample standard deviations given by the methods averaged over the Monte Carlo runs were calculated and compared with the true mean value. The fifth and ninety-fifth percentage points of the between-sample distributions of the estimated mean values were calculated to give some indication of the skewness of these distributions. The true standard deviation of the sample was compared with the average within-sample standard deviation, calculated as the square root of the average over samples of the within-sample variance estimates. Average Monte Carlo estimates of the standard deviation of differences between measured and expected values were also calculated. Correlation coefficients between the estimates for each method and the true values for the individuals were computed and averaged over the Monte Carlo runs, together with between-sample standard deviations, and the minimum and maximum values of the sample correlation coefficients. Estimated distributions of weekly consumption were constructed, on the basis of the 100 simulations of sample size 1,000, again including for comparative purposes the true distribution among the sample members. Monte Carlo estimates of sensitivity and specificity were obtained for thresholds of 500 centilitres of alcohol per year for Finland, and 28 and 50 units per week for England and Wales, values which have been considered as thresholds of heavy drinking in the populations in question (Royal College of Psychiatrists 1979; Royal College of General Practitioners 1986).

The units employed for the amount processes in both simulations correspond to approximate centilitres of ethanol. The Finish beer drinking survey permits fairly accurate estimation of ethanol content of drinks consumed, due to the homogeneous nature of the product. However, the aggregated individual consumption of different alcoholic beverages in the England and Wales survey cannot be expressed very accurately in terms of volume of 100% ethanol. For this reason the results presented maintain a distinction in the terms used to describe the units of measurement.

6. Results

6.1. *England and Wales data*

6.1.1. Sample estimates of population parameters

Table 5 shows the means and standard deviations of estimated weekly amounts by the four measurement methods together with other important summary statistics. To assess the effect of measurement as opposed to sampling error, expected weekly consumption values were calculated for the simulated individuals, and the standard deviations of the differences between the measured values and the expectations are also displayed in Table 5. Note that since the individuals whose drinking histories were simulated in the Monte Carlo runs had the same underlying parameters in each run, the Q-F modal interval method shows no between-run variation.

For all sample sizes the Q-F method based on modal intervals shows negative bias and markedly underestimates the within-sample standard deviation. As would be expected, 24-hour recall exhibits a great deal of variation, both in absolute terms, and in the differences between measured and expected values. Of the three methods based on reporting of actual consumption, 24-hour recall is least accurate, and the differences between last week's consumption and the survey period method are fairly trivial, although the latter shows consistently lower between-sample variation. Q-F based on the eight-week averaging procedure does not appear to suffer from the negative bias

Table 5. Summary statistics from simulation of consumption by males in England and Wales

	Expected consumption	Last week's consumption	Survey period	24-hour recall	Q-F (modal interval)	Q-F (average)
100 simulations of sample size 100						
Mean	19.78	19.75	19.79	19.63	16.93	20.06
s.d.		1.126	1.105	3.157		0.561
5th %ile		17.81	17.86	14.18		19.04
95th %ile		21.87	21.63	25.23		21.03
Within-sample s.d.	17.871	21.160	20.979	35.331	15.724	17.533
s.d. of differences		11.388	11.076	30.242	4.584	6.294
100 simulations of sample size 200						
Mean	19.24	19.27	19.26	19.33	16.82	19.70
s.d.		0.798	0.764	2.170		0.333
5th %ile		17.91	18.01	15.75		9.21
95th %ile		20.60	20.64	23.68		20.24
Within-sample s.d.	17.865	21.188	21.025	35.560	17.333	18.025
s.d. of differences		11.524	11.196	30.826	5.177	6.399
100 simulations of sample size 1,000						
Mean	19.04	19.07	19.06	19.13	16.68	19.22
s.d.		0.366	0.339	0.862		0.165
5th %ile		18.51	18.53	17.53		18.94
95th %ile		19.67	19.61	20.53		19.47
Within-sample s.d.	17.357	20.701	20.481	34.738	16.391	17.172
s.d. of differences		11.233	10.828	29.902	4.887	6.237

associated with Q-F based on modal interval measurement, has the lowest value of between-sample variation, and of all the methods yields the closest estimate to the within-sample standard deviation. The between-sample distributions of the mean estimates appear to be reasonably symmetrical in all cases.

6.1.2. Exceedance proportions

Figure 1 shows the estimated consumption distributions for each of the methods, together with the distribution of expected consumption for the simulated individuals from the 100 Monte Carlo runs of 1,000 simulated respondents. Last week's consumption and survey period methods reproduce the distribution fairly accurately, whereas Q-F and 24-hour recall do not. The downward bias of Q-F based on modal interval measurement is again clear from the position of the curve to the left of the others at higher levels of consumption, while the eight-week average Q-F is closer to the actual distribution. Notice that 24-hour recall overestimates the proportions of consumers below low levels of consumption, and underestimates at high levels. Figure 2

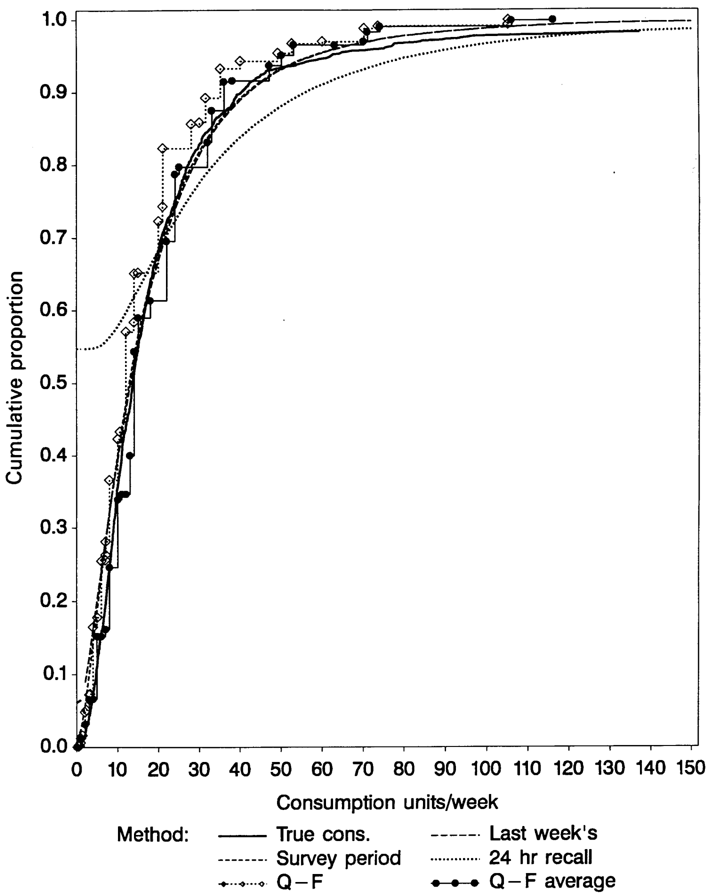

Fig. 1. Cumulative proportions of consumption, England and Wales males 1980

shows the effect of excluding those individuals reporting no consumption in constructing the distributions of consumption. Twenty-four hour recall shows considerable bias of overestimation at all levels, bias in the same direction being evident for last week's consumption method and the survey period method.

Table 6 shows estimated exceedance proportions, sensitivities and specificities together with their standard deviations for the two thresholds. For completeness minima and maxima of the exceedance proportion estimate over simulation runs are also given. The most striking finding is the overestimation of the exceedance proportion by all three methods based on recall of actual occasions, at all sample sizes and both thresholds. The one exception to this is that the survey period method underestimated the exceedance proportion for the 50 units threshold with a sample size of 200. Twenty-four hour recall grossly overestimates at the higher threshold, and even at the lower threshold produces estimates about 25% too large. At both thresholds and for all three sample sizes last week's consumption and survey period measurement perform similarly in terms of sensitivity and specificity, but 24-hour recall is extremely poor. Q-F measurement in general suffers from the categories not converting exactly to the

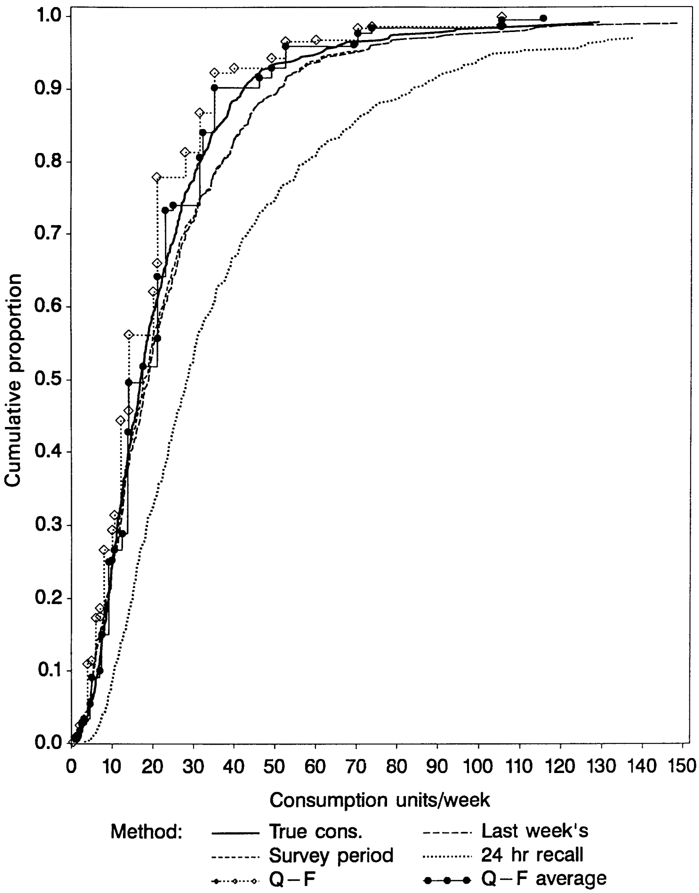

Fig. 2. Cumulative proportions of consumption excluding zeroes, England and Wales males 1980

thresholds. Thus, both versions of Q-F consistently underestimate at the higher threshold, while at the lower threshold Q-F using the averaging process overestimates. The specificity of Q-F is consistently high, again because of the available thresholds, and for the lower threshold the sensitivity of the Q-F modal interval method is comparable with last week's consumption and the survey period methods. For large sample size both forms of Q-F measurement offer good performance on all three indicators related to exceedance proportions. Precision of estimation as measured by the standard deviation of the estimates over simulation runs is similar for last week's consumption and survey period methods, with 24-hour recall markedly inferior for the higher threshold, and less so at the lower. The Q-F average procedure is appreciably more precise than the others.

6.1.3. Correlations with expected values

Table 7 shows the estimated correlation between measured values and expected values expressed as weekly consumption, averaged over the simulation runs, together with other useful summary statistics. Twenty-four hour recall is least satisfactory, both in terms of the magnitude of the average correlation and its associated

Table 6. Estimated exeedance proportions, sensitivities and specificities from simulation of consumption by males in England and Wales

a. Threshold = 50 units/ week

	Last week's consumption	Survey period	24-hour recall	Q-F (modal interval)	Q-F (average)
100 simulations of sample size 100 True value = 0.0600					
Average estimated exceedance proportion	0.081	0.081	0.130	0.030	0.053
(s.d.)	(0.0212)	(0.0212)	(0.0285)		(0.0135)
Minimum estimated e.p.	0.030	0.030	0.060		0.020
Maximum estimated e.p.	0.140	0.140	0.190		0.090
Mean estimated sensitivity	0.727	0.727	0.542	0.500	0.677
(s.d.)	(0.1783)	(0.1765)	(0.1944)		(0.1456)
Mean estimated specificity	0.55687	0.562	0.254	1.000	0.786
(s.d.)	(0.1447)	(0.1451)	(0.0865)		(0.1503)
100 simulations of sample size 200 True value = 0.0800					
Average estimated exceedance proportion	0.080	0.079	0.125	0.050	0.059
(s.d.)	(0.0127)	(0.0121)	(0.0217)		(0.0072)
Minimum estimated e.p.	0.035	0.035	0.070		0.040
Maximum estimated e.p.	0.110	0.105	0.190		0.080
Mean estimated sensitivity	0.674	0.674	0.509	0.563	0.636
(s.d.)	(0.1126)	(0.1126)	(0.1088)		(0.0775)
Mean estimated specificity	0.678	0.685	0.331	0.900	0.870
(s.d.)	(0.0991)	(0.0955)	(0.0685)		(0.0659)
100 simulations of sample size 1,000 True value = 0.0510					
Average estimated exceedance proportion	0.068	0.067	0.119	0.046	0.049
(s.d.)	(0.0058)	(0.0057)	(0.0088)		(0.0032)
Minimum estimated e.p.	0.054	0.054	0.097		0.043
Maximum estimated e.p.	0.089	0.087	0.136		0.057
Mean estimated sensitivity	0.739	0.739	0.544	0.804	0.748
(s.d.)	(0.0633)	(0.0633)	(0.0545)		(0.0484)
Mean estimated specificity	0.553	0.560	0.234	0.891	0.774
(s.d.)	(0.0469)	(0.0464)	(0.0534)		(0.0417)

Table 6. (Continued)

b. Threshold = 28 units/week

	Last week's consumption	Survey period	24-hour recall	Q-F (modal interval)	Q-F (average)
100 simulations of sample size 100 True value = 0.2000					
Average estimated exceedance proportion (s.d.)	0.233 (0.0302)	0.229 (0.0295)	0.251 (0.0346)	0.170	0.244 (0.0196)
Minimum estimated e.p.	0.160	0.150	0.140		0.190
Maximum estimated e.p.	0.300	0.290	0.320		0.300
Mean estimated sensitivity (s.d.)	0.799 (0.0900)	0.799 (0.0900)	0.559 (0.0978)	0.800	0.930 (0.0527)
Mean estimated specificity (s.d.)	0.692 (0.0795)	0.702 (0.0801)	0.449 (0.0730)	0.941	0.764 (0.0472)
100 simulations of sample size 200 True value = 0.1850					
Average estimated exceedance proportion (s.d.)	0.213 (0.0192)	0.209 (0.0198)	0.238 (0.0253)	0.145	0.226 (0.0137)
Minimum estimated e.p.	0.165	0.160	0.190		0.195
Maximum estimated e.p.	0.255	0.255	0.290		0.255
Mean estimated sensitivity (s.d.)	0.770 (0.0633)	0.773 (0.0639)	0.549 (0.0794)	0.757	0.923 (0.0352)
Mean estimated specificity (s.d.)	0.672 (0.0540)	0.686 (0.0562)	0.428 (0.0531)	0.966	0.757 (0.0404)
100 simulations of sample size 1,000 True value = 0.1880					
Average estimated exceedance proportion (s.d.)	0.208 (0.0085)	0.203 (0.0080)	0.240 (0.0114)	0.144	0.202 (0.0054)
Minimum estimated e.p.	0.189	0.187	0.211		0.188
Maximum estimated e.p.	0.229	0.224	0.267		0.219
Mean estimated sensitivity (s.d.)	0.750 (0.0282)	0.751 (0.0287)	0.535 (0.0319)	0.729	0.883 (0.0210)
Mean estimated specificity (s.d.)	0.680 (0.0256)	0.697 (0.0246)	0.420 (0.0224)	0.951	0.821 (0.0182)

Table 7. Correlations between measured and expected values from simulation of consumption by males in England and Wales

	Last week's consumption	Survey period	24-hour recall	Q-F (modal interval)	Q-F (average)
100 simulations of sample size 100					
Average correlation	0.85	0.85	0.51	0.97	0.94
(s.d.)	(0.048)	(0.047)	(0.135)		(0.014)
Minimum correlation	0.71	0.73	0.17		0.89
Maximum correlation	0.94	0.94	0.78		0.96
100 simulations of sample size 200					
Average correlation	0.84	0.85	0.50	0.96	0.94
(s.d.)	(0.030)	(0.031)	(0.079)		(0.009)
Minimum correlation	0.75	0.77	0.29		0.91
Maximum correlation	0.91	0.91	0.69		0.95
100 simulations of sample size 1,000					
Average correlation	0.84	0.85	0.51	0.96	0.94
(s.d.)	(0.015)	(0.015)	(0.043)		(0.005)
Minimum correlation	0.79	0.80	0.41		0.92
Maximum correlation	0.87	0.88	0.62		0.94

variability. As might be expected, sample size has little effect on the magnitude of the correlations, although naturally in larger samples the between-simulation run variation decreases, indicating more precise estimation. Once again there is little to choose between last week's consumption and the survey period methods, but it is notable that the correlation between both Q-F measurements and true values is appreciably higher than the others.

6.2. Finnish data

6.2.1. Sample estimates of population parameters

Table 8 shows summary statistics from the simulation of Finnish beer drinking. Negative bias in Q-F modal interval measurement is apparent at all sample sizes. There is no clear evidence of major bias in any of the other measurement methods. At all sample sizes the survey period method is considerably more precise as measured by the between-sample standard deviation than last week's consumption, which in turn is clearly superior to 24-hour recall, while the Q-F average procedure is overall most precise. The percentiles indicate positive skewness in the between sample distribution of estimated mean consumption for all three recall measures, while the distribution of mean consumption over simulation runs estimated by the Q-F average procedure appears reasonably symmetrical.

In terms of estimation of other aspects, the methods appear to be ordered (best to worst) as Q-F (modal interval), Q-F (average), survey period, last week's consumption and 24-hour recall.

Table 8. Summary statistics from simulation of Finnish beer drinking

	Expected consumption	Last week's consumption	Survey period	24-hour recall	Q-F (modal interval)	Q-F (average)
100 simulations of sample size 100						
Mean	2.60	2.57	2.59	2.42	2.22	2.59
s.d.		0.635	0.395	1.564		0.258
5th %ile		1.63	1.98	0.70		2.20
95th %ile		3.74	3.23	5.57		3.05
Within-sample s.d.	5.106	7.873	6.275	16.306	5.877	5.947
s.d. of differences		5.939	3.639	15.668	2.030	2.696
100 simulations of sample size 200						
Mean	2.618	2.649	2.649	2.726	2.364	2.759
s.d.		0.496	0.290	1.166		0.176
5th %ile		1.99	2.23	0.91		2.48
95th %ile		3.66	3.15	4.66		3.07
Within-sample s.d.	5.262	9.075	6.639	17.778	5.904	6.354
s.d. of differences		6.016	3.862	16.792	1.585	2.627
100 simulations of sample size 1,000						
Mean	2.69	2.68	2.67	2.64	2.70	2.78
s.d.		0.192	0.111	0.527		0.084
5th %ile		2.39	2.50	1.87		2.65
95th %ile		3.05	2.85	3.55		2.91
Within-sample s.d.	5.037	7.796	5.992	16.858	5.772	6.071
s.d. of differences		6.042	3.498	16.092	2.388	2.975

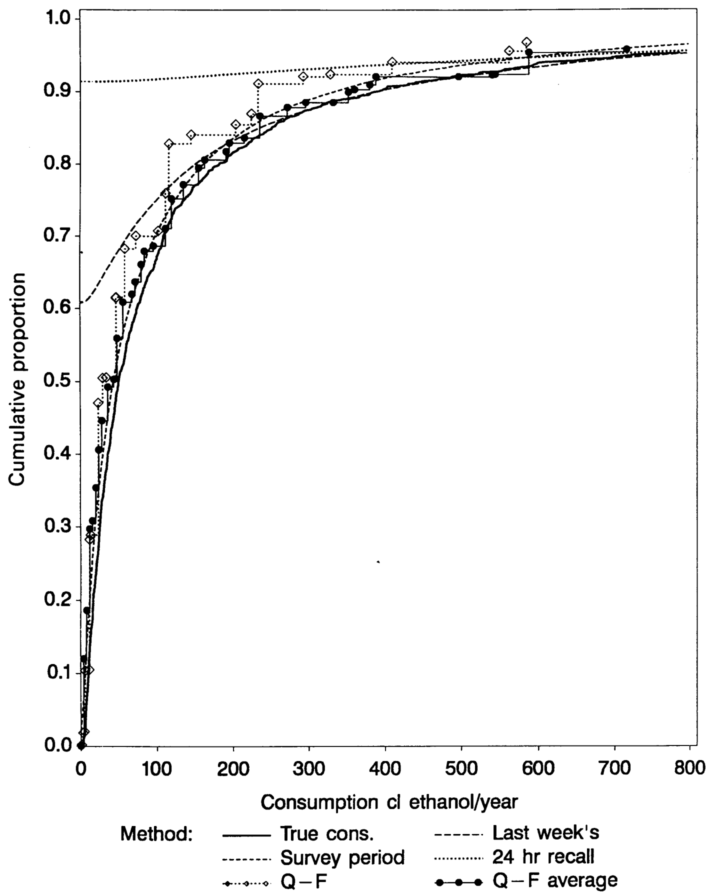

Fig. 3. Cumulative proportions of beer consumption, Finnish males 1984

6.2.2. Exceedance proportions

Figure 3 shows the estimated consumption distributions for each of the methods, together with the distribution of expected consumption for the simulated individuals from the 100 simulations of sample size 1 000. It is clear that the survey period method and the Q-F method assuming an eight-week averaging process reproduce the distribution fairly accurately, whereas the others do not. The problems associated with 24-hour recall in the England and Wales simulation are, as might be expected, even more marked in this population, and are shared to a lesser extent by the last week's consumption method. Once more the downward bias of the Q-F (modal interval approach) is evident. Figure 4 shows that excluding zero consumers introduces extremely large biases in estimation of the distribution for the modal interval Q-F, last week's consumption and 24-hour recall measures. The effect on the survey period method is negligible.

Table 9 shows estimated exceedance proportions, sensitivities, specificities and related statistics for the threshold of 500 cl/year. In terms of exceedance proportion estimation, the bias noted previously in the Q-F (modal interval) method is quite pronounced at the smallest sample size, but decreases with larger numbers of simulated

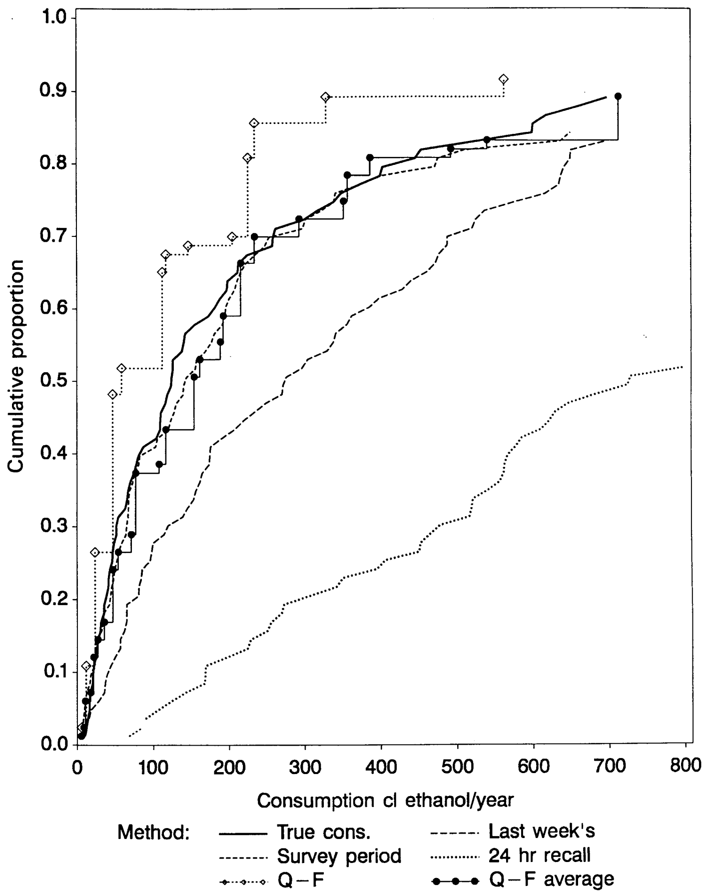

Fig. 4. Cumulative proportions of beer consumption excluding zeroes, Finnish males 1984

respondents. Notwithstanding the result for sample size 200 for the 24-hour recall method, it seems that survey period measurement is the best of the three recall measures in this regard, both in terms of accuracy and precision as measured by the between-samples standard deviation of estimate; it is also clearly superior to the other two in terms of sensitivity and specificity.

Despite the bias mentioned earlier, Q-F interval measurement achieves overall the largest values of sensitivity and specificity, but it is worth noting that if the responses are expressed on the basis of the eight-week averaging process, these indicators are little better than the corresponding survey period values.

6.2.3. Correlations with expected values
Comparison of Table 10 with Table 7 shows that the magnitudes of the correlations in the Finnish beer drinking simulation are in general lower than in the England and Wales consumption simulation, which is a result of the increased variability in the Finnish drinking process. However, in terms of the methods themselves, a similar pattern is evident, and the superiority of the survey period method more marked. Both Q-F measures are again appreciably more highly correlated with expected values.

Table 9. Estimated exceedance proportions, sensitivities and specificities from simulation of Finnish beer drinking

	Last week's consumption	Survey period	24-hour recall	Q-F (modal interval)	Q-F (average)
100 simulations of sample size 100 Threshold = 500 cl/year True value = 0.0600					
Average estimated exceedance proportion	0.070	0.057	0.052	0.040	0.061
(s.d.)	(0.0198)	(0.0143)	(0.0230)		(0.0128)
Minimum estimated e.p.	0.030	0.030	0.010		0.030
Maximum estimated e.p.	0.120	0.090	0.120		0.100
Mean estimated sensitivity	0.653	0.752	0.293	0.667	0.780
(s.d.)	(0.1783)	(0.1700)	(0.1211)		(0.0973)
Mean estimated specificity	0.585	0.812	0.398	1.000	0.792
(s.d.)	(0.1673)	(0.1494)	(0.2445)		(0.1372)
100 simulations of sample size 200 Threshold = 500 cl/year True value = 0.0550					
Average estimated exceedance proportion	0.073	0.060	0.055	0.065	0.078
(s.d.)	(0.0148)	(0.0124)	(0.0174)		(0.0110)
Minimum estimated e.p.	0.040	0.035	0.015		0.050
Maximum estimated e.p.	0.110	0.090	0.095		0.105
Mean estimated sensitivity	0.596	0.688	0.257	1.0000	0.871
(s.d.)	(0.1481)	(0.1319)	(0.1127)		(0.0897)
Mean estimated specificity	0.453	0.688	0.278	0.846	0.624
(s.d.)	(0.0886)	(0.1138)	(0.1408)		(0.0849)
100 simulations of sample size 1,000 Threshold = 500 cl/year True value = 0.0610					
Average estimated exceedance proportion	0.077	0.061	0.056	0.059	0.079
(s.d.)	(0.0071)	(0.0050)	(0.0066)		(0.0050)
Minimum estimated e.p.	0.058	0.050	0.039		0.069
Maximum estimated e.p.	0.095	0.076	0.077		0.091
Mean estimated sensitivity	0.593	0.682	0.239	0.820	0.811
(s.d.)	(0.0625)	(0.0540)	(0.0545)		(0.0426)
Mean estimated specificity	0.474	0.683	0.263	0.848	0.630
(s.d.)	(0.0459)	(0.0474)	(0.0534)		(0.0372)

Table 10. Correlations between measured and expected values from simulation of Finnish beer drinking

	Last week's consumption	Survey period	24-hour recall	Q-F (modal interval)	Q-F (average)
100 simulations of sample size 100					
Average correlation	0.69	0.84	0.34	0.94	0.90
(s.d.)	(0.114)	(0.083)	(0.190)		(0.038)
Minimum correlation	0.31	0.41	−0.04		0.77
Maximum correlation	0.91	0.96	0.69		0.98
100 simulations of sample size 200					
Average correlation	0.68	0.84	0.35	0.96	0.92
(s.d.)	(0.120)	(0.079)	(0.186)		(0.032)
Minimum correlation	0.29	0.53	0.04		0.84
Maximum correlation	0.89	0.94	0.72		0.97
100 simulations of sample size 1,000					
Average correlation	0.64	0.82	0.31	0.91	0.87
(s.d.)	(0.044)	(0.034)	(0.069)		(0.015)
Minimum correlation	0.53	0.71	0.124		0.84
Maximum correlation	0.78	0.88	0.45		0.92

7. Conclusions and Discussion

Of the approaches based on actual occasions, last week's consumption and the survey period methods yield very similar results when applied to the simulated British data, indicating that there is little to be gained from stretching the time period for infrequent consumers when these are a small proportion of the population. In fact if the aim of a survey was simply to estimate mean consumption, with independence between the frequency and amount distributions, and no interest in distribution, exceedance proportions or correlation with true values, it can be shown that it would be better to ask frequent consumers for information about a longer period than infrequent consumers, analogously to the well-known result for stratification by size. This might be impracticable for many reasons and forgetting could be a particular problem, not to mention length of time taken to perform the interview. However the survey period method does show greater specificity in both simulations when used to identify individuals exceeding particular thresholds, and has clear advantages over last week's consumption in the simulation of Finnish beer drinking. The UK simulation shows that for a population of frequent drinkers last week's consumption has no major disadvantages in comparison with the more complicated survey period method. It should be noted however that the practice of excluding drinkers who happened to consume no alcohol during the period in question will bias estimation of the distribution.

Although slightly biased, the Q-F method shows a high correlation with the true parameters of the individuals in both simulations, indicating that it would provide more power than the others in correlational studies of consumption and harm, although regression coefficients would be overestimated. The bias and reduction in variation noted for Q-F measurement are to do with the method of grouping implicit in this method, and it should be noted that if respondents use an averaging process for allocation to Q-F categories the bias is considerably reduced, and in the England and Wales simulation, the corresponding variation is closer to, although still less than, actual. Evidence from existing sample surveys tends to support the notion that individuals report modal behaviour when Q-F questioning is used. For example, Goddard and Ikin (1989) used last week's consumption as their basic method of measurement, and supplemented this with a question regarding the typicality of last week's consumption. About 25% of all respondents reported that last week's consumption was more than usual, while only 8% claimed that they had consumed less than usual in the previous week. If the within-individual distribution of consumption is positively skewed, this result is in accordance with the interpretation of 'typical' by respondents as 'modal' rather than 'average' in the sense of the arithmetic mean.

The results also show that if within-individual variation is great, but individuals categorise themselves into Q-F classes in accordance with either of the algorithms suggested here then Q-F is a more accurate method of measuring mean consumption in terms of measurement variability, although some account must be taken of bias. In connection with the identification of individuals consuming above threshold levels, Q-F is generally superior to the other methods. Once again, the methods of Q-F reporting simulated indicate a degree of negative bias in actual exceedance proportion estimates, but the sensitivities and specificities compare favourably with all the other methods.

Twenty-four hour recall is not a satisfactory method of estimation, except in the case of estimating mean consumption, where it could be improved by increasing sample size, which is the basis of the UK market research approach indicated earlier. However it cannot be expected to reproduce the distribution of consumption and will not be effective as a screening instrument in any population showing considerable temporal variation in consumption. Incorporating a frequency measure for each respondent to weight the reported amounts would improve the performance of this measure.

A recent survey of college students in the United States (O'Hare 1991) compared Q-F and retrospective diary (last week's consumption) approaches on 494 respondents. The retrospective diary approach detected more overall consumption and heavy drinking than Q-F, but the correlation between measured consumption and problems was slightly higher using Q-F measurement than with the retrospective diary approach. These results are in line with the findings of the present study, but it should be clear from the results presented here concerning exceedance proportions, correlations with true values and deviations from true values, that the temptation to prefer methods of measurement which yield higher overall consumption should be resisted.

The main conclusions which may be drawn from this study are that the choice between last week's consumption and the survey period method is of little relevance for a frequently-consuming population although the latter is preferable in populations with a low frequency of consumption, and that 24-hour recall methods are unsuitable for most purposes, especially estimation of the distribution of consumption. Q-F is the best screening method, assuming that individuals categorise themselves in accordance with one of the algorithms suggested here. In terms of directions for further work it seems that the most obvious need is for studies of the way in which individuals summarise their drinking behaviours in terms of Q-F categories, and a useful first step would be a study of the relationship between longterm drinking histories and Q-F measures in a real population.

In this study no attempt was made to model forgetting or concealment, but future work may incorporate forgetting processes of the types suggested by Mäkelä (1971), Sikkel (1985) and Sikkel and Jelierse (1988). In fact, as seen above, the structure of the simulation program makes it easy to include measurement errors, such as the probability of forgetting occasions as a function of time and errors in reporting the amounts. Furthermore, the flexibility of the program allows the use of probability models other than the ones reported here; for instance non-hormogeneous Poisson processes, intervals between occasions other than the exponential, and so on. As more is learned about alcohol consumption behaviour, future work along these lines will become both more necessary and more practicable.

8. References

Alanko, T. (1984). An Overview of Techniques and Problems in the Measurement of Alcohol Consumption. In Research Advances in Alcohol and Drug Problems, Vol. VIII, edited by R.G. Smart et al., 209–226, New York: Plenum Press.

Breeze, E. (1985). Women and Drinking. London: HMSO.

Central Statistical Office of Finland (1985). Kotitaloustiedustelu 1985. (Consumer Expenditure Survey 1985). Helsinki: Central Statistical Office of Finland.

Devroy, L. (1986). Non-uniform Random Variate Generation. New York: Springer-Verlag.

Duffy, J.C. and Waterton, J.J. (1984). Randomised Response Models for Estimating the Distribution Function of a Quantitative Character. International Statistical Review, 52, 165–171.

Duffy, J.C. (1986). The Distribution of Alcohol Consumption – 30 Years On. British Journal of Addiction, 81, 735–741.

Ekholm, A. (1968). A Study of the Drinking Rhythm of Finnish Males. 28th International Congress on Alcohol and Addiction. Washington: ICAA.

Fishman, G.S. (1978). Principles of Discrete Event Simulation. New York: John Wiley.

Goddard, E. (1991). Drinking in England and Wales in the Late 1980s. London: HMSO.

Goddard, E. and Ikin, C. (1989). Drinking in England and Wales in 1987. London: HMSO.

Goodhardt, G.J., Ehrenberg, A.S.C., and Chatfield, C. (1984). The Dirichlet: A Comprehensive Model of Buying Behaviour. Journal of the Royal Statistical Society, Ser. A, 147, 621–655.

Gordon, T., Kagan, A., Garcia-Palmieri, M., Kannel, W.B., Zuckel, W.J., Tillotson, J., Sorlie, P., and Hjortland, M. (1981). Diet and its Relation to Coronary Heart Disease and Death in Three Populations. Circulation, 63, 500–515.

Johnson, N.L. and Kotz, S. (1970a). Distributions in Statistics: Continuous Univariate Distributions – I. New York: John Wiley.

Johnson, N.L. and Kotz, S. (1970b). Distributions in Statistics: Continuous Univariate Distributions – II. New York: John Wiley.

Ledermann, S. (1956). Alcool, Alcoolisme, Alcoolisation. Paris: Presses Universitaires de France.

Lemmens, P. and Knibbe, R. (1991). Seasonal Variation in Survey and Sales Estimates of Alcohol Consumption. In Measurement and Distribution of Alcohol Consumption, edited by P. Lemmens. Maastricht: Rijksuniversiteit Limburg.

Mäkelä, K. (1978). Measuring the Consumption of Alcohol in the 1968–1969 Alcohol Consumption Study. Helsinki: Social Research Institute of Alcohol Studies.

Michael, J., Schucany, W., and Haas, R. (1976). Generating Random Variables Using Transformations with Multiple Roots. The American Statistician, 30, 88–90.

Midanik, L. (1982). The Validity of Self-reported Alcohol Consumption and Alcohol Problems: A Literature Review. British Journal of Addiction, 77, 357–382.

O'Hare, T. (1991). Measuring Alcohol Consumption: A Comparison of the Retrospective Diary and the Quantity—Frequency Methods in a College Survey. Journal of Studies on Alcohol, 52, 500–502.

Office of Population Censuses and Surveys (1990a). Family Expenditure Survey Annual Report, 1988. London: HMSO.

Office of Population Censuses and Surveys (1990b). General Household Survey, 1988 Report. London: HMSO.

Poikolainen, K. and Kärkkäinen, P. (1983). Diary Gives More Accurate Information about Alcohol Consumption than Questionnaire. Drug and Alcohol Dependence, 11, 209–216.

Poikolainen, K. and Kärkkäinen, P. (1985). Nature of Questionnaire Options Affects Estimates of Alcohol Intake. Journal of Studies on Alcohol, 46, 219–222.

Room, R. (1985). Measuring Alcohol Consumption in the U.S.: Methods and Rationales. Paper presented at the ICAA-Epidemiology Section Meeting, Rome.

Royal College of General Practitioners (1986). Alcohol – A Balanced View. London: Royal College of General Practitioners.

Royal College of Psychiatrists (1979). Alcohol and Alcoholism. London: Tavistock.

Sichel, H.S. (1982). Repeat-buying and the Generalized Inverse Gaussian-Poisson Distribution. Applied Statistics, 31, 193–204.

Sikkel, D. (1985). Models for Memory Effects. Journal of the American Statistical Association, 80, 835–841.

Sikkel, D, and Jelierse, G. (1988). Renewal Theory and Retrospective Questions. Applied Statistics, 37, 412–420.

Simpura, J. (1981). Scandinavian Drinking Survey: Construction of Indices of Alcohol Intake. Oslo: National Institute for Alcohol Research.

Simpura, J. (ed) (1987). Finnish Drinking Habits. Results from Interview Surveys in 1968, 1976 and 1984. Helsinki: Finnish Foundation for Alcohol Studies.

Sudman, S. and Bradburn, N.M. (1982). Asking Questions. San Francisco: Jossey-Bass.

Waterton, J.J. and Duffy, J.C. (1984). A Comparison of Computer Interviewing Techniques and Traditional Methods in the Collection of Self-report Alcohol Consumption Data in a Field Survey. International Statistical Review, 52, 173–182.

Wilson, P. (1980). Drinking in England and Wales. London: HMSO.

Wilson, P. (1981). Improving the Reliability of Drinking Surveys. The Statistician, 30, 159–167.

Received April 1991
Revised June 1992