

Methods of estimating individual levels of alcohol consumption in the general population

Margaret Ely, Rebecca Hardy, Nicholas T. Longford, Michael E.J. Wadsworth

Introduction

Knowledge of individual levels of alcohol consumption in the general population is derived from surveys in which a sample of subjects from the population is asked what they drink. The estimates derived from the data collected from these general population surveys invariably fall short of estimates of national consumption based on excise duty. This is termed the problem of 'coverage' and generally survey estimates represent only about 40-60% of annual sales (Pernanen, 1974). In other words self-reported alcohol consumption in general population surveys appears to dramatically underestimate consumption. There are two factors which are usually cited to account for the underestimation:

- (i) under-reporting and
- (ii) low response rates in general population drinking surveys.

Neither of these would present a problem if the resulting estimates were not biased. Underreporting could be adjusted for if all responders underreport to a similar degree. However, it is widely assumed that under-reporting affects heavier drinkers more than light drinkers because they are more likely to forget how much they have drunk or because they are more likely to deliberately under-report their drinking from self-consciousness. In other words, self-reported consumption gives biased estimates of actual consumption level.

Although low response rates in low power, the estimates themselves are not biased if non-response is not related to drinking - that is, those who do not respond drink at similar levels to those who respond to the survey. However, it is also believed that heavy drinkers are more likely to refuse to take part in surveys or are less likely to be contacted because they change address more frequently, or are homeless (Lemmens et al., 1988). If this is the case, alcohol consumption is systematically different for non-responders and responders and hence estimates based on the sample of responders are not representative those of the general population.

This report summarises findings from the MR. NSHD on alcohol consumption at the age of 43. It examines how the estimates of alcohol consumption are affected by the survey instrument used and by non-response, and the implications of these findings for the collection of alcohol consumption data in general population surveys.

2. Methods

2.1 Subjects

The MR. National Survey of Health and Development (NSHD) is a follow-up of legitimate, single births to all wives of non-manual and agricultural workers and to one in four wives of manual workers in England, Wales or Scotland during the week 3rd-9th March 1946, a sample of 5362 births. A wide range of information on social, psychological, educational, medical and biological characteristics of the study members has been collected on nineteen occasions from infancy, childhood and adult life (Wadsworth, 1991). At the last of these occasions, in 1989, trained nurses interviewed the study members when they were 43 years old. At this time, 3262 (85%) of the 3854 with whom contact was attempted were interviewed; 4 (0.1%) had died 11 (0.3%) were living abroad, 106 (2.7%) were permanent refusals, 195 (5.1%) temporarily refused because of personal or family problems and 276 (7.2%) could not be traced. Of the 1508 of the birth sample whom there was no attempt to contact 361 (24%) had died, 607 (40%) were living abroad and 540 (36%) had permanently refused to take part at a previous contact (Wadsworth et al, 1992). Excluding study members who were living abroad, whom the survey did not intend to represent, and those who had died, 74.5% (3262/4379) of those in the birth cohort who were still alive and resident in England, Wales or Scotland were interviewed at the age of 43. These respondents are in most respects representative of the native population born in the early postwar years (Wadsworth et al., 1992).

Apart from response rates themselves, the results reported have been weighted to adjust for the initial sampling procedure.

2.2 Outcome Measures : alcohol consumption and drink problems

The information about alcohol was collected at the end of the 1989 interview in two distinct ways: firstly using a self-completion questionnaire and secondly using a diet diary.

The self-completion questionnaire contained questions about alcohol consumption during the past seven days (7 day recall) and also the CAGE questionnaire (Ewing, 1984).

CAGE

The four CAGE questions, each with yes/no response options are as follows:

Have you ever felt you ought to Cut down on your drinking? (Do not include dieting.)
Have people ever Annoyed you by criticising your drinking?
Have you ever felt bad or Guilty about your drinking?
Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hangover? (Eye-opener)

The questions were asked of lifetime experience of such problems and where the answer to a question was 'yes', study members were asked whether they had experienced this 'in the last year'.

The CAGE score is defined as the number of affirmative answers to these questions, ranging from 0 to 4; there are two scores, one relating to lifetime experience of problems (CAGE-EVER) and one relating to problems in the last year (CAGE-LASTYEAR).

7 Day Recall

Consumption of alcohol was based on responses to the question 'in the last seven days how many of the following drinks have you had?' Three categories of drink were differentiated:

spirits (measures of spirits or liqueurs); wine (glasses of wine, sherry, martini or port) and beer (half pints of beer, lager, cider or stout). Thus the quantities reported were approximately equivalent to units of alcohol.¹

7 Day Diet Diary

The 7 day diet diary was used to record all food and drink, including alcohol, consumed during each day of one week. The diary was begun at the end of the interview when the research nurse interviewer recorded all food and drink consumed in the previous two days, demonstrating the method and detail required. The study member was then asked to keep the diary for the subsequent five days and to return it by post in the pre-paid envelope supplied. A carbon copy of the first two days was retained by the nurse so that at least two days of dietary information was available from those who were interviewed. Thus the first two days of the diary are a retrospective record completed by the interviewer prompting the study member for information, the subsequent five days are a prospective record kept by self completion by the study member. The period to which the diary refers overlaps with that referred to by the 7 day recall since the first two days of the diary refers to the last two days of the week recalled in the summary measure.

The diary comprised daily sheets each identified by date and the day of the week, providing spaces on which to record meals, including alcoholic and non-alcoholic beverages, and between-meal snacks, ending with a reminder section to record any other snacks or drinks not previously recorded. This reminder section included specific prompts for beer, wine, sherry or spirits besides such items as sweets, tea, and other common items of food which may not have been taken with a meal. The layout of the sheets was structured in three columns headed: 'food/drink', 'description and preparation' and 'amount'. The weight and nutritional composition of all food and drink recorded in the diaries was derived by the Dunn Nutrition Unit, Cambridge using a computerised system, DIDO (Diet In Data Out) together with a suite of programs for nutritional analysis (Price et al., 1995). The alcoholic content of drinks was converted to grams per 100ml. This conversion was based on a study of the average alcohol content of 29 types of beers, ciders, wines, liquors and spirits derived from samples of each type (Paul and Southgate, 1978). The total grams of alcohol consumed per day was calculated from the quantities and types of drink reported. The data used in the analysis below consists of 7 items per subject, each item being the alcohol consumption on each of the 7 days of the diary. Where appropriate, grams of alcohol were converted to units, the measure of alcohol consumption generally used in Britain, by dividing by 7.9 (Royal College of Physicians, 1996).

Categories of drinker

It is not obvious how we should measure the extent of 'under-reporting' in the diary relative to the recall. One measure is the difference between the total weekly consumption given by the two measures. Using this measure underreporting increases with consumption level because people who drink a lot tend to have larger differences. For example an extra 10 units in the diary total is more likely, and less 'significant', for someone who drinks around 100 units in a week compared with someone who only drinks around 20 units. Although the additive difference in reported consumption is the same in each example, the first case represents a proportional underreporting of only 10%, compared with 50% in the latter case. If proportional differences are used the opposite is true: proportional differences in the two measures decrease with consumption: people who drink little tend to have large proportional differences. For example someone who reports drinking nothing on recall but reports

¹ A unit is roughly equivalent to half a pint (290 ml) of ordinary (4 per cent) beer or lager, 1 pub measure (24 ml) of spirit, 1 glass (50 ml) of sherry or fortified wine or 1 glass (125 ml) of wine.' Guidelines for Health Promotion Number 46, Faculty of Public Health Medicine, Royal Colleges of Physicians of the United Kingdom.

drinking only 1 glass of wine in the diary has multiplied their drinking by a factor of 8, the same as someone who reports drinking 10 pints of beer in recall and 80 pints in the diary. Measuring underreporting by neither additive nor proportional differences is satisfactory. In the light of public health information on the implications of levels of drinking, the solution adopted here is to base comparisons on the classification of drinkers as 'sensible', 'immoderate' or 'heavy' according to the accepted criteria (Royal College of Psychiatrists, 1986), as shown in Table 1. The respondents are classified according to their reported consumption in both the recall and diary measures. Agreement between the two instruments is measured by the agreement in these classifications. The proportion of respondents who are in a lower category in their recall than in their diary is used as a measure of underreporting in the recall relative to the diary.

Table 1: Classification of drinkers according to reported consumption levels

description	level of weekly consumption in Units	
	men	women
sensible	0 to 21	0 to 14
immoderate	over 21, up to 50	over 14, up to 35
heavy	over 50	over 35

Results

Under-reporting in weekly recall compared with daily diary

Comparison of the results from the two instruments used for collecting information about alcohol consumption in the MRC NSHD provides evidence of under-reporting in the recall measure. The average total alcohol consumption reported in a weekly period based on the recall is 9.5 Units (weighted n=2214) while the diary gives a higher average of 13.4 Units (weighted n=1356). There are greater differences in the estimated proportions over given limits of consumption using these two measures, as shown in Table 2. These are significantly higher in the diary than in the recall, except for heavy drinking women (over 35 Units), of whom there are very few cases in either the recall (8) or the diary (4).

Table 2: Estimated proportions drinking above recommended limits using recall and completed diaries

limit (U)	7 Day Recall		7 Day Diary (completers)	
	%	95% CI	%	95% CI
women				
14	6.1	(4.7, 7.6)	15.4	(12.7, 18.1)
35	0.7	(0.2, 1.2)	0.6	(0.01, 1.1)
n (weighted)	1092		694	
men				
21	21.0	(18.6, 23.4)	35.4	(31.8, 39.0)
50	4.0	(2.9, 5.2)	10.0	(7.7, 12.3)
n (weighted)	1123		661	

We cannot expect that a subject would report the same consumption in the recalled week and diary week because the two measures do not relate to the same period. The comparison of the two measures will be affected by variation in drinking pattern from week to week, although this will be reduced since the periods covered by the 7 day recall and the diary overlap by two days. In fact, agreement between the two measures arises only for those who declared no consumption in both instruments (16% of the sample; 10% of men and 21.7% of women). These subjects are likely to drink only infrequently or never at all. If the difference between the recall and diary totals is explained by the variation in levels of drinking from week to week one would expect there to be an equal proportion of respondents with positive and negative values for the difference in diary and recall consumption. But of the respondents who recorded both their 7 day recall and completed the diary (weighted n=1327), only 24.1% declared more on the recall than on the diary, whilst 59.9% declared more in the diary than in the recall².

The instruments were according to their classification of respondents as sensible, immoderate or heavy drinkers (as defined in Table 1). The results for the men and women who completed both the diary and the recall are given in Table 3. Only 2.6% (17/651) of men and 1.0% (7/677) of women were classified in a lower drinking category using their diary than they were using their recalled drinking; whereas 22.9% (149/651) of men and 10.5% (71/677) of women were classified in a higher drinking category in the diary than in the recall. Most of the latter were classified as sensible drinkers by their recall, but as

² The difference, 35.8%, is statistically significant. The 95% confidence interval for the difference is (32.3, 39.3).

immoderate drinkers by their diary total (20% of men and 11% of women who recalled drinking at a 'sensible' level). In addition almost one third (34/109) of male drinkers classified as immoderate by their recall were classified as heavy drinkers by their diary declarations.

Table 3 : Classification of respondents' drinking level according to Recall and Diary total .
(Weighted analysis of the men and women who completed both the diary and the recall).

3.1 Men

column % (n)		Recall			all (%)
		sensible	immoderate	heavy	
Diary	sensible	77.8(403)	13.8(15)	0.0(0)	64.2
	immoderate	20.3(105)	55.0(60)	8.3(2)	25.7
	heavy	1.9(10)	31.2(34)	91.7(22)	10.1
	all (%)	79.6	16.7	3.7	n=651

3.2 Women

column % (n)		Recall			all (%)
		sensible	immoderate	heavy	
Diary	sensible	89.1(566)	13.2(5)	0.0(0)	84.3
	immoderate	10.7(68)	81.6(31)	50.0(2)	14.9
	heavy	0.2(1)	5.3(2)	50.0(2)	0.7
	all (%)	93.8	5.6	0.6	n=677

In conclusion, the differences between reported consumption in the diary and in the recall could not be accounted for by variations in drinking from week to week. The reported consumption in the diary was generally in excess of that reported in the recall measure, and the differences have a substantial effect on estimates of excessive drinking. For two reasons the differences in reported drinking in the diary compared to the recall can be viewed as a measure of the extent to which the recall measure underestimates actual drinking. Firstly, the higher estimates given by the diary are in closer agreement with estimates from sales. Secondly, subjects are unlikely to have had any motivation to over report their drinking in the diary as has been reported for alcoholics in clinical treatment (personal communication, Professor John B. Davies, Centre for Applied Social Psychology, University of Strathclyde). The proportion of the sample who were classified as a higher category of drinker in their diary than in their recall measure is a measure of the extent to which the recall measure underestimates alcohol consumption.

Psychological attitudes to drinking and differences between reported consumption using the two instruments

Although the recall measure underestimates consumption this could be adjusted for if the extent of under-reporting was similar for all respondents. However, we found that the extent of under-reporting was associated with respondents' attitudes to drinking. Those who reported having problems with drinking in the CAGE questions were more likely to under-report their consumption.

The CAGE questionnaire was developed as a screening device for alcoholism in clinical settings, a score of two or more being used as an indicator of risk. More recently it has been increasingly used in general population surveys in Britain, in which a score of two or more is used to indicate drink problems (Hedges, 1997; Richards et al, 1997; Hope et al, 1998). We used responses to the CAGE questions concerning problems in the previous year to indicate current problems with drinking. Apart from the question concerning use of an 'eye-opener' the CAGE questions reflect psychological attitudes to drinking. Comparison of

underestimation in those who responded affirmatively compared to those who did not was used to indicate the sensitivity of the instrument of measurement to attitudes to drinking.

The results showing the specific changes in category according to recall and diary reported consumption are given in Tables 4 - 8 in Appendix 1. These are summarised in Table 9. Firstly, we compared those with drink problems, defined as a CAGE score of two or more, with others. The results shown in Table 4 are summarised by the proportion who underestimated their drinking in the recall relative to the diary given in Table 9. Thirty six percent of men with problems compared to 21% of other men underestimated their drinking in the recall compared to the diary, whilst the corresponding proportions for women were 41% and 9%. Both men and women who reported having drink problems were more likely than those who did not to underestimate their drinking. This analysis was repeated with the individual CAGE questions: 'cut down' (Table 5), 'annoy' (Table 6), 'guilty' (Table 7) and 'eyeopener' (Table 8), and the results are summarised in Table 9.

Table 9 Proportions underestimating consumption in recall relative to diary by gender and CAGE. (Weighted analysis of the men and women who completed both the diary and the recall).

	n	women		n	men	
		%	95% CI		%	95% CI
CAGE - (<2)	643	8.9	(6.7 11.1)	571	21.2	(17.8 24.5)
CAGE + (>1)	34	41.2	(24.6 57.7)	78	35.9	(25.6 46.5)
<i>Cut down</i> - no	619	7.8	(5.6 9.9)	505	17.6	(14.3 20.9)
<i>Cut down</i> - yes	56	39.3	(26.5 52.1)	145	42.1	(34.0 50.1)
<i>Annoy</i> - no	662	10.1	(7.8 12.4)	591	21.0	(17.7 24.3)
<i>Annoy</i> - yes	13	23.1	(0.2 46.0)	57	43.9	(31.0 56.7)
<i>Guilty</i> - no	641	8.6	(6.4 10.7)	584	21.6	(18.2 24.9)
<i>Guilty</i> - yes	34	41.2	(24.6 57.7)	63	33.3	(21.7 45.0)
<i>Eyeopener</i> - no	673	10.4	(8.1 12.7)	639	23.3	(20.0 26.6)
<i>Eyeopener</i> - yes	3	0.0	(0.0 0.0)	10	0.0	(0.0 0.0)

Respondents who had felt they ought to cut down on their drinking (in the past year) were more likely to underestimate their drinking using recall than those who had not felt this way. This was true of men and women alike. Men who reported having been annoyed by criticisms of their drinking were more likely to underestimate their drinking in the recall (43.9%) than those who did not report feeling this way (21.0%). Results for the women were in the same direction - 23.1% vs. 10.1% - but these results were inconclusive because too few women (n=13) reported feeling annoyed at the criticisms of others. Women who had felt guilty about their drinking were more likely to underestimate their drinking in the recall than those who had not felt this way, but this was not so apparent in men. Interestingly responses to the eye-opener question, the one question which was related to physical symptoms rather than psychological attitude, did not seem to be associated with changes in categories of drinker. However the numbers of men (n=10) and women (n=3) who responded affirmatively to this question were too small to support any conclusion.

The results can be summarised as follows. The tendency to underestimate consumption in recall relative to diary is greater for those who admit to problems with drinking measured by

the CAGE questions. Both men and women's reporting in the recall relative to the diary was affected by their feeling that they ought to cut down on their drinking. The recall instrument is sensitive to feelings of guilt about drinking in women and to feelings of annoyance at criticism by others in men.

Non-response

Types of non-response

In many sample surveys some of the people selected to be in the sample do not take part at all, for example because they cannot be contacted or they refuse to participate (case non-response); others do not respond to all the items in the survey (item non-response). In the NSHD those study members who were not interviewed at age 43 give rise to case non-response, and those who were interviewed at age 43 but did not complete the drink questions give rise to *item* non-response. In this report we are concerned with non-response by those who were interviewed at age 43 to the items asking for 7 day recall of drinking, CAGE questions and completion of a weekly diet diary. Each of these items consist of several parts, so that non-response may be partial (for example failure to complete the number of beers in the last week whilst wines and spirits were completed), or complete (for example completing neither beers, wines nor spirits).

Extent of non-response

The extent of complete non-response was similar for each of the items in the 7 day recall, CAGE-LASTYEAR, and diet diary, and accounted for a small proportion of respondents (2.7%, 2.1%, and 2.3%, respectively). However complete non-response to CAGE-EVER was greater. Six percent of the respondents answered none of the questions relating to life time experience of problems with drinking. This difference between CAGE-EVER and CAGE-LASTYEAR might be explained by the layout of the questions. The questions relating to the last year were on the right hand side of the paper and could have been more visible to the respondent (see Appendix 2). Respondents who omitted one of the three items generally responded to another, so that some information regarding drinking was available for all but 0.4% (13 cases) of the respondents at age 43 (see Figure 1). In only 14 cases was there was no information regarding quantity of alcohol consumed.

There were greater differences in the extent of partial non-response. Besides those who did not answer any of the questions in the items, an additional 0.8% omitted at least one of the CAGE-LASTYEAR questions, 21.9% omitted at least one of the types of drink in the recall, while 36.3% omitted to record at least one of the designated diet diary days (see Figure 2) Although the extent of partially missing data in the items used to measure alcohol consumption presents a more serious problem, in these cases information about the respondents could be obtained from the parts that had been completed or from information provided by the alternative measure. Most respondents provided some information for both the recall and the diet diary items (see Figure 1).

Influences on non-response which are unrelated to drinking behaviour

In order to assess the likely impact of missing values on the estimates produced by a particular survey we first consider evidence relating to the specific circumstances of the survey. In this survey all the items referring to alcohol came at the end of a long interview and in some cases the nurse was unable to complete the survey because of pressure of time. The recall and CAGE items were asked at the end of a self-complete questionnaire which was itself at the end of a long³ structured interview. The diet diary instructions and recording of the first two days of the diary constituted the final part of the interview. The remaining five days of the diet diary were to be completed prospectively by the respondent who had to forward them in an envelope provided, without further reminders. While this survey methodology ensured a high level of response to the first two diary days (96.6%), it also explains the high level of partial non-response since 30% of the respondents failed to return any further diary sheets (see Figure 2). There may be many reasons which are

³ The mean length of the interview was 2 hours 12 minutes and 5% of interviews were more than two and a half hours long.

unrelated to the drinking behaviour for the failure to return the diary sheets. Since the diet diary was not designed to collect information specifically about alcohol consumption and non-response to this item consisted of a failure to complete sheets on general dietary consumption, it might well be that diary non-response was not related to the alcohol consumption of the respondent. We can, however, use the rich collateral information about the respondents contained in this survey to estimate missing data. The method for doing this, multiple imputation, is explained below. For specific questions on drinking such as the recall and CAGE items, we may have more reason to suspect that non-response may be related to consumption.

Non response and levels of drinking

A plausible explanation of non-response to the specific questions on alcohol contained in the recall and CAGE items is that respondents who did not drink disregarded the questions altogether. In this survey there was no identification of non-drinkers and questions about recalled consumption and problems were directed at all respondents (see Appendix 2). Women are generally more conscientious in answering survey questions than men (ref? - awaiting response from SCPR). For example in this study women were more likely than men to complete a diet sheet, yet they were less likely than men to answer CAGE-EVER or recall questions. One hypothesis is that women are more likely to be non drinkers and to find the questions irrelevant.

All the 128 respondents who left the CAGE 'ever' questions blank, but who answered the 'last year' questions, denied having problems in the last year. Of these, all who answered questions about recalled consumption in the previous week reported drinking no alcohol. Anecdotal evidence was provided by the research nurse interviewers that study members who did not drink skipped the CAGE questions because they saw them as irrelevant.

While partial missing responses to the diet diary may be expected because of the commitment that such an instrument required of respondents, the high level of partial non-response to the concise recall questions called for some other explanation. There is evidence that subjects left the question blank instead of entering a zero when they had not drunk, or perhaps because they never drank, that particular type of alcoholic drink. Although 25% (806) of the respondents left at least one of the recall questions blank, only 0.6% (19) used both blank and a zero. In addition, the distribution of blanks by gender and social class was similar to that of the zero entries. For example, women were more likely than men to leave their beer consumption blank, particularly women from higher social classes, and these were exactly the groups who were likely to drink no beer. The reverse was true for wine, which is drunk more by women than by men, and by those in higher social classes.

Psychological attitudes to drinking and nonresponse to alcohol consumption questions

Attitudes to drinking represented by responses to the CAGE questions were not associated with non-response to the recall questions. *All but six (91%) of the respondents who did not respond to the CAGE also omitted the recall questions as well (see Figure 1).* The proportions not completing the diary by gender and CAGE are given in Table 11. The attitudes of men towards their drinking were not associated with non-completion of the diary. In fact a lower proportion of men who reported drink problems in the past year failed to complete their diaries than those who did not report such problems.(40.6% vs. 42.1%) The reverse was true of women (44.1% vs. 37.9%), although the difference was not large enough to reach statistical significance at the 5% level. The proportion of women not completing their diaries was particularly high for those who had felt annoyed by other people

criticising their drinking (58.8%), and for those who reported having a drink first thing in the morning (63.6%). However numbers in the latter group were too small to reach statistical significance. This is in contrast with the relationship of the specific CAGE questions with underreporting, for which women's feelings of guilt and men's feelings of annoyance seemed more important.

Table 11 Proportions not completing the diary by gender and CAGE.

	n	women		n	men	
		%	95% CI		%	95% CI
CAGE - (<2)	1045	37.9	(35.0 40.8)	993	42.1	(39.0 45.2)
CAGE + (>1)	59	44.1	(31.4 56.7)	133	40.6	(32.3 48.9)
Cut down - no	999	37.4	(34.4 40.4)	871	41.6	(38.3 44.8)
Cut down - yes	100	45.0	(35.2 54.8)	251	43.0	36.9 49.2
Annoy - no	1066	37.4	(34.5 40.3)	1032	42.3	(39.3 45.4)
Annoy - yes	34	58.8	(42.3 75.4)	92	37.0	(27.1 46.8)
Guilty - no	1049	38.3	(35.4 41.3)	1013	42.1	(39.0 45.1)
Guilty - yes	52	34.6	(21.7 47.5)	107	41.1	(31.8 50.4)
Eyeopener - no	1092	37.9	(35.0 40.8)	1106	42.0	(39.0 44.9)
Eyeopener - yes	11	63.6	(35.2 92.1)	18	38.9	(16.4 61.4)

In general reported problems with drinking were not associated with non response as they were with underreporting consumption. In particular the relationships between the specific CAGE questions and these two outcomes were quite different.

Multiple imputation of missing values of alcohol consumption

Omitting incomplete records for alcohol consumption results in reduced statistical power and, when non response is related to alcohol consumption, in biased estimates. The alternative is to enter or 'impute' a likely value in place of each missing item (such as a zero for those who omitted one of the recall questions). If only one such value is imputed for each missing value, it will be treated in any analysis as if it were an actually recorded value, and no account will be taken of our uncertainty about the value that would have been recorded had the subject answered the question. Only multiple values can reflect such uncertainty. One complete data set is constructed using each of the imputed values. The required analysis is then applied to each complete data set and the results are averaged (with standard errors being adjusted for sampling variation within and between the data sets).

We have substantial collateral information about the missing items of alcohol consumption in the diary. To impute a missing diary value we take into account background factors which are relevant to their alcohol consumption, for example: gender, smoking status, body mass, response to the CAGE questions, recorded alcohol consumption in recall and in previously recorded diary days, and the day of the week. We assume that the missing alcohol item is similar to that reported by respondents with the same background values, by selecting a number of values from the distribution of the alcohol consumption amongst these respondents.

Results obtained from multiple imputation of diary values

Using multiple imputation there was no difference in the estimates of the proportions drinking above the sensible and heavy drinking limits (see Table 12). The principal gain in the use of multiple imputation in this survey is in the reduction in the standard errors for the estimates. Such a reduction is achieved by the use of the information in the incomplete records. This result is subject to the model of missing items which assumed that non completers were like diary completers in the relationship between the alcohol consumption and the background variables.

Table 12: Estimated proportions drinking above recommended limits using those who completed the diary and the whole sample of respondents who were interviewed in 1989, using multiple imputation of missing diary items.

limit (U)	Diary completers		Diary All interviewed using multiple imputation	
	%	se	%	se
women				
14	15.4	1.37	14.2	1.12
21	5.9	0.89	5.7	0.71
35	0.6	0.29	0.9	0.32
n	694		1131	
men				
21	35.4	1.86	33.2	1.47
28	25.3	1.69	23.7	1.33
50	10.0	1.17	9.1	0.93
n	661		1147	

Multiple imputation for missing items in the recall changed the estimates of recall totals only slightly (less than 1% change drink category as a result of imputing values rather than replacing blanks with zeros).

Implications for surveys of alcohol consumption

It has been suggested that in assessing measures of alcohol consumption, the closer the measure comes to assessing actual quantities consumed, the more reliable the measure. The less the measure requires the subject to reflect on or summarise their drinking, the less bias in the reported consumption (Alanko, 1981). Our results support this hypothesis. They show that the diary gives greater coverage than the recall, a result supported by studies in the Netherlands (Lemmens et al., 1992). We have also been able to examine the relationship of underreporting in the recall to attitudes to drinking measured by the CAGE. Our results show that reporting in the diary is less biased by attitudes to drinking than in the recall. These results support the conclusion that estimates based on the diary are more valid than those based on the recall. However, the more complex instrument is more expensive to collect and analyse and suffers from greater item non-response. Asking about alcohol consumption using more than one instrument provides information which can be used to estimate missing data. Simple summary measures like the recall provide background information for the estimation of missing data in more complex instruments.

The collection of the first two days of the diary by nurses retrospectively and the retention of a duplicate copy of this part of the diary by the nurse ensures that partial information is available for most of the respondents. This information can be used to estimate consumption on the rest of the week for those respondents who do not complete or return the diary. The previous days' drinking were the most important contributors to the estimation of alcohol consumption in the diary days that were not completed. Collecting just two days retrospective diary data saves on interviewer time and is therefore cheaper than collecting a full 7 days retrospectively, the method used in specialist general population surveys conducted in the UK by the OPCS (Dight, 1976; Breeze, 1985; Goddard, 1989). The method of multiple imputation provides a way of estimating the missing data that results from the failure of respondents to return their full diary. The combination of the methods of collection of two days retrospective data together with multiple imputation therefore provide a solution to the collection of alcohol consumption data in non specialist surveys with limited resources for the collection of alcohol data.

Our results do not support the usual assumption that non response is more prevalent amongst heavy drinkers, using the data collected at age 43. It seems that item non-response, at least in non-specialist surveys, may not be related to higher levels of consumption. In fact, item non-response in recall is associated with lower levels of consumption, or non-drinking of a particular beverage. The evidence from the multiple imputation results suggests that non response to the diary is not related to the level of alcohol consumption. This conclusion is subject to the assumption that non-responders are like responders with respect to the relationship between their alcohol consumption and the background information.

Self reported problems with drinking affected the sensitivity of the instrument used to measure alcohol consumption more than the extent of non-response. This would suggest that problem drinkers are more likely to underestimate their drinking than to omit questions about alcohol consumption. In this respect the diary yields more valid estimates of consumption than a summary measure.

Non-specialist surveys are likely to collect more valid data on alcohol consumption than specialist surveys because estimates of consumption are less likely to be biased by non-response. Those who do not complete the survey are less likely to do so because of reasons connected with their drinking when the survey is not specifically directed at the investigation of alcohol consumption.

Missing data can be estimated using background information and partial responses. Multiple imputation provide a simple way of reflecting the uncertainty in the estimates due to missing values. It uses all the information in the incomplete records and hence reduces the standard errors of the estimates. It therefore increases the statistical power of any further analyses of alcohol consumption compared with those which use only cases where complete data has been recorded. The lack of impact of the multiple imputation on estimates is specific to this particular survey. If non completion had been greater amongst people who were like those who reported heavy drinking in the survey, then this would have been reflected in higher estimates. This situation could arise, for example, in a specialist drink survey using a diary which was specifically about drinking behaviour. Also, although the multiple imputation have little impact on the estimates of alcohol consumption, this does not preclude the possibility that the relationships between risk factors and alcohol consumption are affected by their use. Even though non-response to the diary may not be associated with, say, marital status, nor with excessive consumption, the relationship between marital status and excessive consumption amongst non responders may be different from that amongst responders. Thus inferences concerning the relationships between risk factors and excessive drinking based on the whole of the interviewed sample using multiple imputation may be different from such inferences based on the sample of diary completers only.

There are two aspects of measurement which have not been dealt with in this report. The first is that the diary, by providing daily estimates of alcohol consumption, provides information about patterns of consumption which are not available in the recall of a weekly total. Patterns such as daily excess, or number of drinking days, are more relevant than a weekly total to the recent Department of Health guidelines on sensible drinking (Department of Health, 1995). Secondly, we have not dealt with the problem of non-response to the survey at age 43. We do not know how much alcohol was consumed by those people who had dropped out of the survey by that age. However, the methods used here could be extended to investigate this question.

References

- Alanko, T. (1981) An overview of techniques and problems in the measurement of alcohol consumption. In R.G. Smart et al., (eds) *Research advances in alcohol and drug problems* 8, 209-226. New York: Plenum Press.
- Breeze, E. (1985) *Women and Drinking*. OPCS Social Survey Division. HMSO, London
- Department of Health (1995) *Sensible Drinking: the report of an interdepartmental working group*. Department of Health, London.
- Dight, S. E. (1976) *Scottish Drinking Habits*. OPCS Social Survey Division. HMSO, London.
- Ewing, J.A. (1984) Detecting alcoholism, the CAGE questionnaire. *Journal of the American Medical Association* **252**, 1905-07.
- Goddard, E. (1989) *Drinking in England and Wales in the late 1980s*. OPCS Social Survey Division. HMSO, London.
- Hedges, B. (1997) Alcohol Consumption. In *Health Survey for England, 1994.*, Colhoun, H. and Prescott-Clarke, P. eds, pp. 338-365.
- Hope, S., Power, C. & Rodgers, B. (1998) The relationship between parental separation in childhood and problem drinking in adulthood. *Addiction* **93**, 505-514.
- Lemmens, P., Tan E. S. and Knibbe, R. A. (1988) Bias due to nonresponse in a Dutch Survey on Alcohol Consumption. *British Journal of Addiction* **83**, 1069-1077.
- Lemmens, P., Tan E. S. and Knibbe, R. A. (1992) Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. *Journal of Studies on Alcohol* **53**, 476-486.
- Pernanen, K (1974). Validity of survey data on alcohol use. In Gibbins et al (eds) *Research advances in alcohol and drug abuse*. Vol 1. Wiley, 1974
- Price, G.M., Paul A.A., Key, F.B., Harter, A.C., Cole, T. J., Day, K.C. and Wadsworth M.E J. (1995) Measurement of diet in a large national survey: comparison Of computerised and manual coding of records in household measures. *Journal of Human Nutrition and Dietetics* **8**, 417-428.
- Paul, A. A. and Southgate, D. A. T. (1978) McCance and Widdowson's *The Composition of Foods* (4th ed.). London: HMSO.
- Richards, M., Hardy, R. and Wadsworth, M. E. J. (1997) The effects of divorce and separation on mental health in a national birth cohort. *Psychological Medicine* **27**, 1121-1128.
- Royal College of Physicians (1996) *Preventing the harm related to alcohol use; reducing population risk*. Guidelines for health promotion number 46. Policy Office, Faculty of Public Health Medicine, London.
- Royal College of Psychiatrists (1986) *Alcohol: Our Favourite Drug* .Tavistock Publications, London and New York.
- Wadsworth, M. E. J. (1991) *The Imprint of Time : Childhood, History and Adult Life*. Clarendon press, Oxford
- Wadsworth, M. E. J., Mann, S. L., Rodgers, B., Kuh, D. J. L., Hilder, W. S., Yusuf, E.J. (1992) Loss and representativeness in a 43 year follow up of a national birth cohort. *Journal of Epidemiology and Community Health* **46**, 300-304.