INTRODUCTION

Recently interest in standardized methods for assessment of persons with potential or fully developed alcohol problems has increased in the Swedish health care and social service systems. This is partly due to increased needs for cost-effectiveness and standardized information bases for documentation and treatment planning. There are a number of assessment methods in this area which, however, can be difficult to choose from (Allen et al., 1995). Thus, one has to be clear about the purpose of the assessment. We have found a 4-step conceptual assessment model valuable in this regard (screening for identification, problem assessment, personal assessment and follow-up assessment).

To address alcohol problems, it is an advantage to identify persons with hazardous or harmful alcohol use and if possible intervene before severe alcohol problems develop. To identify persons with severe alcohol problems, different versions of the Michigan Alcoholism Screening Test (MAST) (Selzer, 1971) have often been used. Screening methods for hazardous and harmful drinking are, e.g., CAGE (Cut-down, Annoyed, Guilt, Eye-opener; Ewing, 1984) and AUDIT (Alcohol Use Disorders Identification Test, Saunders et al., 1993). AUDIT is a quick screening method used in many countries to identify persons with potential or established drinking problems, particularly in primary care settings. Since AUDIT is sensitive, not only to severe alcohol problems, but also to hazardous drinking, it is particularly suitable for studies of the general population, where the prevalence of alcohol problems is lower than in clinical samples. It is recommended by the World Health Organization (WHO). In a review (Allen et al., 1997), AUDIT has been proposed as a psychological alcohol marker which in medical contexts might be combined with biological alcohol markers, e.g., γ-glutamyltransferase (GGT), mean corpuscular volume (MCV) and carbohydrate-deficient transferrin (CDT). Both the validity in terms of sensitivity and specificity, and the internal reliability of AUDIT are generally reported to be satisfactory. However, the test–retest reliability is seldom evaluated.

AUDIT consists of 10 items with five response alternatives (the last two items only three), scored from 0 to 4 points (see Table 1). It is usually administered as a self-report questionnaire, but can also be given as interview or via computer. The items are sampled from three content domains: ‘alcohol consumption’ (items 1–3), ‘signs of alcohol dependence’ (items 4–6) and ‘alcohol-related harm’ (items 7–10). The maximum score is 40. A ‘standard drink’ is defined as 12 g of 100% alcohol/day (± 20%). The Swedish test form has been designed for optical reading and computer scoring. It has recently become very popular both in research and routine use.

It is of conceptual and practical value to evaluate the factor structure of AUDIT in samples with varying prevalence of alcohol problems. In a study of a non-clinical, ‘low alcoholism prevalence’ sample in Mexico consisting of 2050 male workers, Medina-Mora et al. (1998) performed a principal components analysis and reported a two-factor structure. One factor comprised the consumption items 1 to 3. Items 4 to 10 constituted an ‘alcohol problems’ factor. However, the tenth item loaded above 0.40 on both factors. The same factors were found by Maisto et al. (2000) in a Canadian sample of 7035 male and female primary care patients. As in the Mexican study the tenth item was split on both factors.

In a study of 100 psychiatric emergency patients (Bergman et al., 1998), the validity of AUDIT for predicting alcohol dependence or alcohol misuse according to DSM-III-R diagnoses (American Psychiatric Association, 1987) was found to be satisfactory. The point-biserial correlation between the diagnoses and AUDIT score was 0.71 (P = 0.000). The validity of a screening tool like AUDIT depends on the chosen cut-off score and type of sample investigated (i.e. prevalence of alcohol problems). Thus, in this high prevalence sample (46% of the patients later acquired alcoholism diagnoses), a cut-off score of ≥15 was found to give the highest positive prediction value, much higher than the recommended score of ≥8 for general use. The internal consistency reliability was similar for men and women (Cronbach alpha = 0.95). The reliability of the three a priori, defined subscales corresponding
to the three content domains was also satisfactory: 0.88 (hazardous alcohol use), 0.93 (alcohol dependence) and 0.85 (harmful use). The high reliabilities have been taken as credit for the possibility of presenting subscale results besides AUDIT total score. However, a principal components analysis revealed one single factor explaining ≥70% of the variance. This result is consistent with an AUDIT study of drug abusers (Skipsey et al., 1997) also reporting a single factor explaining a substantial proportion of the variance. In a study of 2000 persons (92% men) suspected of drunken driving, i.e. again a ‘high alcohol problem prevalence sample’, Bergman et al. (2000) clearly showed that principal axis factoring of the AUDIT inter-item correlation matrix resulted in only one factor explaining 51% of the variance. All factor loadings were 0.49 or higher. A two-factor solution was also explored but this led to splitting. In sum, in one Mexican and one Canadian ‘low alcohol problem prevalence sample’, two factors emerge corresponding to ‘hazardous consumption’ and ‘alcohol-related problems’, whereas in one Swedish and one American ‘high alcohol problem prevalence sample’, the results speak in favour of a single factor.

In order to evaluate a person’s AUDIT results and to give more meaningful feedback to a respondent, reference values from the general population of corresponding age and sex should be useful. However, to our knowledge, only one population AUDIT survey has been carried out. This was done by Holmila (1995) in Finland. She reported the proportion of positive cases identified and also (as expected) that harmful consequences of drinking were more frequent in persons drinking often and in large quantities. Since the reliability was not reported, the psychometric quality of the Finnish AUDIT translation is unknown and, unfortunately, Holmila (1995) did not report results in such a way that they can be used as reference values for general use. A study by Díaz (cited in Medina-Mora et al., 1998) surveyed the poor population in a central area of Mexico. However, besides not being representative of the general population, useful reference values were also not reported in this study. Lacking reference values from the general population, a person’s result in AUDIT is generally compared with a cut-off score, most often ≥8 (hazardous or harmful alcohol use, Saunders et al., 1993). For severe alcohol problems, a cut-off score of ≥19 has been proposed (Claussen and Aasland, 1993). However, partly dependent on the purpose of the testing, cut-off scores varying between 2 and 15 points have been used (Allen et al., 1997). The present study is the first Swedish general population alcohol survey using AUDIT. As men usually score higher on AUDIT than women, the question of whether the cut-off score for hazardous or harmful alcohol use should be set lower for women than for men becomes an important one. There are two arguments for a lower female cut-off score. First, women develop a higher blood-alcohol level than men after the same amount of alcohol consumed per kg body weight (Bradley et al., 1998a). A mean difference of 22% has been reported (Källmén, 1995). Secondly, the risks for medical alcohol-related harm, e.g. liver cirrhosis and cognitive disorder, are higher for women than for men (Bergman and Hindmarsh, 1987; Bradley et al., 1998a). Because of that, a 25% lower limit for acceptable alcohol intake has been suggested for women (Sanchez-Craig et al., 1995). In two recent reviews (Bradley et al., 1998b; Damström-Thakker, 1998) of the health hazards of drinking, it was suggested that women should drink no more than one ‘standard drink’ (10–14 g of absolute alcohol) and men two (20–28 g) a day. Since binge drinking refers to the drinking of many drinks on the same occasion in order to get drunk, the operationalization in AUDIT is six drinks or more. It has been suggested that the definition of binge drinking in AUDIT should be changed to four, instead of six, drinks on the same occasion for women (Bradley et al., 1998b). Keeping the binge drinking definition at six drinks, a 25% lower cut-off score would imply that women who score ≥6 should be identified as positive cases.

The present study had four purposes: (1) the main purpose was to investigate alcohol use among Swedes as assessed by AUDIT and provide age- and gender-specific reference values; (2) a related purpose was to try out the effects of a 25% lower cut-off score in AUDIT for women on the prevalence of hazardous or harmful alcohol use; (3) the factor structure of AUDIT was also to be investigated. The hypothesis, derived from the Medina-Mora et al. (1998) and Maisto et al. (2000) studies that the AUDIT inter-item correlations can be explained by two factors, rather than only one, as suggested by the Skipsey et al. (1997) and Bergman et al. (1998) ‘high alcohol problem prevalence samples’, or perhaps even by three as implied from the item content domains, was also to be tested; (4) the psychometric quality in terms of internal and test–retest reliability of AUDIT and its subscales based on the results of the factor analysis was also evaluated.

**SUBJECTS AND METHODS**

**Sample and drop-out**

A random sample of 1250 persons, half men and half women, aged 17–71 years, was drawn from an official register of the addresses of all persons living in Sweden. The sample...
Statistical considerations

We took advantage of this to compute positively skewed, i.e., most respondents obtain low and a few particularly in 'low-prevalence samples' is not normal but positively skewed; i.e., most respondents obtain low and a few obtain high scores. For such distributions, it is often recommended to use non-parametric instead of parametric statistics and analysis procedures or use logarithmically transformed values instead of raw scores. However, we have chosen parametric statistics and procedures on raw scores to make our results comparable to previous AUDIT studies. Furthermore, the factor analyses performed on both raw scores and logarithmically transformed values resulted in nearly identical factor matrices. It is a general experience that parametric and non-parametric statistical analyses most often give the same overall results.

RESULTS

As expected, men scored higher than women (5.0 ± 4.7 vs 2.7 ± 2.8) on the AUDIT total score. With the usual cut-off score of ≥28, 17.9% of the men and 5.1% of the women were identified as having hazardous or harmful alcohol use. Only 14 men (3%) and three women (1%) scored ≥19, which indicates that there were few persons with severe alcohol problems among the respondents. The proportion of women who reported not using alcohol at all (18.1%) was almost twice as high as the corresponding male proportion (9.8%). When the cut-off score for women was decreased to 26, the number of positive cases increased to 10.6%.

With a class width of 11 years, the sample was divided into five age groups: 17–27 (n = 184), 28–38 (n = 219), 39–49 (n = 230), 50–60 (n = 218) and 61–71 years (n = 146). The AUDIT score decreased with increasing age. The 17–27-year-old men scored highest (7.1 ± 6.1) and the women aged 61–71 years lowest (1.4 ± 1.3). An analysis of the responses to item 2 (How many ‘glasses’ do you drink on a typical day when you drink alcohol?) revealed that men in the age group 17–27 years on average drank 4–5 glasses on each occasion and the 61–71-year-old men 1–2 glasses. The 17–27-year-old women reported an average of 2–3 glasses per occasion while the 61–71-year-old women drank merely 1 glass. Item 1 (How frequently do you drink alcohol?) did not reveal the same strong age dependence, nor did the binge drinking item (How frequently do you drink six such glasses or more on the same occasion?), where all age groups of both sexes responded on average with ‘never’ or ‘more rarely than once a month’.

The AUDIT inter-item and item-total product-moment correlations are presented in Table 2. The item-total correlations varied between 0.36 and 0.71. Thus, frequency of binge drinking explained a lot more of the AUDIT total score variance (52%) than frequency of drinking (13%). A confirmatory factor analysis (CFA, Jöreskog and Sörbom, 1993), testing the hypothesis that the inter-item correlation matrix can be explained by two factors, resulted in acceptable (Marsh et al., 1988; Bollen, 1990) adjusted goodness of fit (AGFI = 0.94) and normed and non-normed fit indices (NFI = 0.95 and NNFI = 0.94) despite a χ² value of 189.02 with 34 df (P = 0.000). The alternative hypotheses of one or three factors were also evaluated. The one-factor solution was discarded due to much lower fit indices but the three-factor solution resulted in the same fit indices as the two-factor solution and a lower χ² (166.12 with 32 df; P = 0.000). To facilitate the choice between the two or three factor models, an exploratory principal axis factor analysis with Varimax rotation was also carried out. According to the ‘eigenvalue ≥1.0’ and the ‘scree plot’ criteria, two factors explaining 55% of the variance were found. The three consumption items corresponding to ‘hazardous consumption’ constituted the first and the seven items from the
‘dependency’ and ‘alcohol-related harm’ content domains, henceforth called ‘alcohol-related problems’, constituted the second factor (Table 3). Extracting a third factor resulted in item splitting. Thus, items 2 and 3 loaded on both the first and the second factor and items 8, 9 and 10 loaded on both the second and the third factor. The three-factor model was discarded in favour of the more parsimonious two-factor model. The communality of item 9 was somewhat lower than for the other items. The same results were achieved when the factor analyses were carried out on log-transformed item scores.

The product-moment correlation between the AUDIT subscales (all items given the same weight) based on the factors ‘hazardous consumption’ and ‘alcohol-related problems’ was 0.57 ($P = 0.000$) and between total score and the subscales ‘hazardous consumption’ and ‘alcohol-related problems’ 0.88 ($P = 0.000$) and 0.89 ($P = 0.000$), respectively.

The Cronbach alpha coefficient of AUDIT total score was 0.82, somewhat lower among women (0.75) than among men (0.83), and the coefficients for the subscales ‘hazardous use’ 0.69 and ‘alcohol-related problems’ 0.80. The test–retest reliability was 0.98 for ‘alcohol-related problems’ and 0.93 for total score and ‘hazardous use’. Thus, there was a high response stability (intra-class correlation) across the 3–4-week interval.

The main effects of gender and age and their possible interaction on the AUDIT scores were tested by two-way ANOVA. Gender and age were independent variables and AUDIT score the dependent. Significant effects of gender and age ($P = 0.000$) but no interaction ($P = 0.80–0.99$) were found in AUDIT total score and in the two subscales (see Table 4). This means that the relationship between age and AUDIT score was similar for both genders and the two sexes were brought together in the next phase of the analyses.

In order to check which of the age classes could be collapsed conveniently, pairwise $F$-tests according to Scheffé’s ex post facto method were performed. The results clearly indicated that the five age groups could be collapsed into three: 17–27 years (97 men and 87 women), 28–60 years (307 men and 360 women) and 61–71 years (67 men and 79 women). Means and SD of each age group and gender are shown in Table 5 and can be used to compute normative standard scores, such as $T$-scores (50 ± 10). When transforming a raw score to the corresponding non-normalized $T$-score, the tabulated mean value of the appropriate gender and age group is subtracted from the observed raw score and divided by the tabulated SD. The result of this calculation is multiplied by 10 and finally 50 is added. Such $T$-scores for the three age groups by gender based on the responses of our population sample are shown in Table 6.
Table 5. Means, SD and proportion of positive cases for total score, ‘hazardous consumption’ and ‘alcohol-related problems’ by age group and gender

<table>
<thead>
<tr>
<th></th>
<th>Men (years)</th>
<th>Women (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17–27</td>
<td>28–60</td>
</tr>
<tr>
<td>Total score</td>
<td>4.7 ± 4.1</td>
<td>2.5 ± 2.4</td>
</tr>
<tr>
<td>Problems</td>
<td>3.4 ± 2.0</td>
<td>2.1 ± 1.6</td>
</tr>
<tr>
<td>28 (%)</td>
<td>1.7 ± 3.0</td>
<td>0.4 ± 1.3</td>
</tr>
<tr>
<td>26 (%)</td>
<td>19.5</td>
<td>2.8</td>
</tr>
</tbody>
</table>

DISCUSSION

In this general population study, a satisfactory response rate of ~80% was acquired. Men had a somewhat lower response rate than women. Since a comparison between the AUDIT results of early and late responders showed that they did not differ significantly, it was assumed that the non-responders probably did not bias the results of the study to a degree that would seriously invalidate our conclusions. However, we probably lost some heavy drinkers, especially 39–49-year-old men and 17–27-year-old women, leading to an underestimation of the prevalence of hazardous alcohol use and alcohol-related problems, particularly in these age groups. To correct for this has not been possible, due to the anonymity of the responders.

Since the female prevalence of hazardous and harmful alcohol use was largest among 17–27-year-old women, the underestimation is probably of most importance in this group.

Item 3 concerning frequency of binge drinking (six or more glasses on single occasions) explained half of the AUDIT score variance. A similar result was found in the Finnish (Holmila, 1995) and Mexican (Medina-Mora et al., 1998) studies. This supports the suggestion that the binge drinking question is the best item indicator of hazardous or harmful alcohol use of the entire test in non-clinical samples (Bush et al., 1998). In many studies not using AUDIT, binge drinking is defined as five drinks or more for men and four or more for women (Wechsler et al., 1995). However, changing the definition of binge drinking by a lower number of drinks in AUDIT might lower the correlation between this item and total score, and, furthermore, render AUDIT result comparisons between studies more difficult. In the Mexican, but not in the present Swedish, sample the item concerning frequency of alcohol consumption had approximately the same high correlation with total score as the binge drinking item. This observation suggests that frequency of consumption is a better predictor of hazardous or harmful alcohol use in Mexico, than in Sweden, a somewhat unexpected result since the drinking culture in both countries is mainly built upon liquor, rather than wine. In our clinical sample of psychiatric emergency patients and drivers suspected of drunken driving, it was item 7 instead (How frequently did you have feelings of guilt or a bad conscience from your alcohol consumption?) that had the highest correlation with the AUDIT total score. This difference might be due to the fact that the proportion of persons with alcohol problems and with good reason to have a bad conscience for heavy drinking, is much higher in these populations than in the general population.

Drinking habits of women have changed during the last 30 years (Bengtsson et al., 1998). Moderate, but not heavy, drinking has increased since the 1960s, consistent with a convergence between male and female drinking habits (Neve et al., 1996). However, there is still a big sex difference in this regard. In the present population survey, about every tenth man and every fifth woman was a teetotaller, whilst 17.9% of the men and 5.1% of the women had hazardous or harmful...
alcohol use according to the commonly used cut-off score of ≥8 in AUDIT. Since women are recommended a lower daily intake due to higher sensitivity for the acute and chronic effects of alcohol, a 25% lower female cut-off score was tested, i.e. ≥6 points instead of ≥8. The proportion of female positive cases then identified doubled, from every twentieth, to every tenth, woman.

The prevalence of hazardous or harmful alcohol use according to AUDIT ≥8 has been shown in four different male samples to vary between 13.4 and 17.9% (Bergman et al., 1998; Diaz cited in Medina-Mora et al., 1998; Hermansson et al., 2000; present study). The corresponding proportion among women varied much more, between 0.7 and 8%, probably due to cultural differences, e.g. level of female emancipation, between Mexico and Sweden.

In order to compare the proportion of positive AUDIT cases in our population survey with that of Holmila’s (1995) Finnish study, the cut-off score was increased to ≥11 points. The proportion of positive Swedish cases was lower than the proportion of positive Finnish ones; 10 versus 22% among men and 1.9 versus 5% among women. The discrepancy reflects different drinking cultures and registered alcohol retailing. In Finland, the alcohol retailing in litres of 100% alcohol per resident in 1996 was nearly 7 l and in Sweden 5 l (Folkhälsoinstitutet and CAN, 1998). The discrepancy should be evident on both ‘hazardous alcohol consumption’ and ‘alcohol-related problems’ due to the significant correlation between the two AUDIT subscales implied by Holmila’s (1995) study and reported in the present study. Hazardous or harmful drinking among women seems to be as prevalent in Australia (5% scored ≥11, Fleming, 1996) as in Finland.

As reported in previous studies using AUDIT (Fleming et al., 1991) and not using AUDIT (Wechsler et al., 1995), our young respondents drank much more alcohol than older respondents. However, not only ‘hazardous consumption’ scores but also ‘alcohol-related problems’ scores and the proportion of positive cases identified decreased with increasing age. Thus, ≥40% among men aged 17–27 years as compared with 15% among the 28–60-year-old group and 8% among the 61–71-year-old men scored ≥8. The corresponding prevalence of women scoring ≥6 was 26, 9 and 0%. According to Table 6, a raw score of 8 points corresponds to a T-score of 52, 58 and 65 respectively among the three male age groups. A raw score of 6 for a woman corresponds to a T-score of 53, 65 and 85 depending upon age. This trend reflects more liberal norms of alcohol use, more risk-taking behaviour and greater needs for stimulation and novelty among young persons (Nezlek et al., 1994). Drinking decreases markedly between 21 and 28 years of age. Thus, the decrease in AUDIT scores from 17–27 to 28–38 years in our sample was striking. The maturing process is probably due to family building and the fact that many young women are informed that alcohol and pregnancy do not go together. Those at risk for continued or escalated drinking are most frequently men also showing other problem behaviours (Bennett et al., 1999). However, before recommending a higher cut-off score for young people, more research about what factors interact to develop alcohol problems later in life is needed.

The factor structure of AUDIT in the present population study was nearly identical to that of the Medina-Mora et al. (1998) Mexican study. Thus, two factors emerged: ‘hazardous consumption’ (items 1, 2 and 3) and ‘alcohol-related problems’ (items 4–6 indicating alcohol dependence and items 7–10 indicating alcohol-related harm). The factor structure was computed, not only on raw scores, but also on logarithmically transformed values to avoid the violation of the non-normality of the raw score item distributions. The results were nearly identical. This supports the robustness of the factor structure. By contrast, when factor-analysing AUDIT in clinical samples with a high prevalence of alcohol problems, only one factor seems to emerge due to higher and more equal inter-item correlations between consumption, signs of dependence and alcohol-related harm. Non-normalized age- and gender-corrected T-scores corresponding to raw scores on the two subscales based on the factors ‘hazardous consumption’ and ‘alcohol-related problems’ can be calculated from Table 4 and are presented in the Swedish AUDIT manual together with percentile scores (available from the first author). Age- and gender-corrected reference values should be useful in research for controlling the effects of these two crucial factors and also in routine use when giving feedback of AUDIT results to a respondent.

To make correct decisions based on assessment methods, such as AUDIT, the method must be reliable and valid. The Swedish version of AUDIT was shown to have satisfactory internal and test-retest reliability. AUDIT is not a diagnostic tool but a screening method to identify persons with hazardous or harmful alcohol use according to a total score based on all 10 items. However, based on our analyses, a more qualitative subscale approach in terms of ‘hazardous consumption’ and ‘alcohol-related problems’ can be a valuable addition to the presentation of AUDIT total score. A modestly elevated total score generally depends on hazardous consumption, whereas a pronounced elevation, particularly a score of ≥19, also indicates alcohol-related problems including dependence.

Further research on alcohol use in Sweden, as assessed by AUDIT, is planned to take place in the years 2001 and 2005 in order to evaluate the effects from more liberal alcohol import quotas and lowered alcohol taxes due to the Swedish membership in the European Union.

Acknowledgements — This study was supported by The Swedish Retailing Monopoly’s foundation for alcohol research (SFA) and has been approved by the Ethical Committee of the Karolinska Institute.

REFERENCES


ALCOHOL USE AMONG SWEDES


