Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption—II

JOHN B SAUNDERS,1 OLAF G. AASLAND,2 THOMAS F. BABOR,3 JUAN R. DE LA FUENTE4 & MARCUS GRANT5

1Centre for Drug and Alcohol Studies, Royal Prince Alfred Hospital, and Departments of Medicine and Psychiatry, University of Sydney, Sydney, Australia, 2Ministry of Health and Social Services, Oslo, Norway, 3Alcohol Research Center, Department of Psychiatry, University of Connecticut Health Center, Farmington, Connecticut, USA, 4Instituto Mexicano de Psiquiatría, Mexico City, Mexico & 5Programme on Substance Abuse and Mental Health Division, World Health Organization, Geneva, Switzerland

Abstract
The Alcohol Use Disorders Identification Test (AUDIT) has been developed from a six-country WHO collaborative project as a screening instrument for hazardous and harmful alcohol consumption. It is a 10-item questionnaire which covers the domains of alcohol consumption, drinking behaviour, and alcohol-related problems. Questions were selected from a 150-item assessment schedule (which was administered to 1888 persons attending representative primary health care facilities) on the basis of their representativeness for these conceptual domains and their perceived usefulness for intervention. Responses to each question are scored from 0 to 4, giving a maximum possible score of 40. Among those diagnosed as having hazardous or harmful alcohol use, 92% had an AUDIT score of 8 or more, and 94% of those with non-hazardous consumption had a score of less than 8. AUDIT provides a simple method of early detection of hazardous and harmful alcohol use in primary health care settings and is the first instrument of its type to be derived on the basis of a cross-national study.

Introduction
Early intervention is a new approach to the prevention and management of alcohol-related problems. It is a pro-active technique which aims to identify persons with hazardous or harmful alcohol consumption* before dependence and serious harm have occurred, and to provide brief therapy, typically at the point of first contact (World Health Organization, 1980; Babor, Ritzon & Hodgson, 1986; Saunders, 1987; Institute of Medicine, 1990; Saunders & Foulds, 1992).
The primary goal of intervention is to facilitate reduction in alcohol intake to non-hazardous levels, and thereby lessen the risk of harmful consequences of drinking. There is increasing evidence for its efficacy in this regard (Kristenson et al., 1983; Wallace, Cutler & Haines, 1988; Institute of Medicine, 1990; Saunders & Foulds, 1992). Through early identification, treatment is provided well before the stage alcohol-affected patients usually present for treatment, which typically is when they have suffered major health problems, psychological impairment and social decline. Early intervention is particularly appropriate for primary health care settings, where the high throughput of patients offers great opportunities for systematic screening and therapy.

An essential component of any early intervention procedure is a simple and valid screening instrument to detect persons with hazardous or harmful alcohol consumption before dependence and permanent harm have developed. The present paper describes the derivation of a 10-item screening questionnaire which is based upon experience in a World Health Organization collaborative study based in six countries (see Appendix). It follows a companion paper (Saunders et al., 1993) which describes the prevalence of alcohol use disorders and the inter-relationships of consumption and harm in a primary health care population. The instrument, the Alcohol Use Disorders Identification Test (AUDIT) is a development of the provisional 'core' screening instrument described in the original report of the collaborative project (Saunders & Aasland, 1987).

Screening instruments for alcohol use disorders are not new. Nearly all existing instruments have, however, been developed to detect alcoholism, and not to screen for problem drinking which is at an earlier or milder stage. The most widely known alcoholism questionnaire is probably the Michigan Alcoholism Screening Test (MAST) (Selzer, 1971), of which several versions now exist (Pokorny, Miller & Kaplan, 1972; Selzer, Vinokur & Van Rooijen, 1975; Swenson & Morse, 1975). Questions were selected for the MAST on the basis of their capacity to discriminate between inpatient alcoholics and psychiatric patients who had no diagnosis of alcoholism. The MAST correctly identified 98% of alcoholics and only 5% of non-alcoholic patients had a positive score. Several items in the MAST reflect advanced physical dependence or severe harm, or whether the subject identifies as an alcoholic. Examples include "Have you ever had delirium tremens (DT's)?" and "Have you ever attended a meeting of Alcoholics Anonymous?". Among the broader spectrum of problem drinkers, the MAST is less sensitive: 59% of persons convicted of drunken and disorderly behaviour and only 11% of persons whose vehicle licences were under review because of an alcohol-related offence were identified as 'cases' (Selzer, 1971). In a community study, Saunders & Kershaw found that the short version of the MAST and a related instrument, the CAGE, identified only a minority of problem drinkers (Saunders & Kershaw, 1980). One may conclude that these instruments are useful and very sensitive in screening for advanced problems such as alcoholism, but are less suitable for detecting those with less severe drinking problems, who actually form a larger proportion of the general population (Mayfield, McLeod & Hall, 1974; Saunders & Kershaw, 1980; Bernadt et al., 1982; Jacobson, 1983; Ewing, 1984; Hedlund & Vieweg, 1984; Bush et al., 1987; Waterson & Murray-Lyon, 1989). Similar comments may be applied to several other screening instruments such as the Munich Alcoholism Test and the MacAndrew Alcoholism Scale (MacAndrew, 1965; Jacobson, 1983; Feuerlein et al., 1986; Gottesmann & Prescott, 1989).

In Europe, screening techniques for alcoholism have been developed which are based on the presence of abnormal physical findings. The principal exponent of this approach has been the French physician Le Go. The 'Le Go Grid' (Le Go, 1976) is a brief clinical screening procedure which involves the detection of, inter alia, conjunctival injection, abnormal vascularization of the facial skin, coating of the tongue and hepatomegaly. It has been taken up widely in France for the detection of alcoholism (Babor et al., 1988). To date there has been little work on the use of the Le Go Grid in detecting less severe drinking problems.

Biological markers of heavy alcohol consumption have also been used to screen for problem drinking. They include serum gammaglutamyltransferase (GGT) activity, serum aspartate and alanine aminotransferase activities (AST and ALT respectively), HDL-cholesterol, uric acid and erythrocyte mean cell volume (MCV)
The sensitivity of these markers is variable: up to 80% of alcoholics admitted to general medical or gastroenterology units have a raised GGT or MCV (Rosalki & Rau, 1972; Wu et al., 1974). However, in psychiatric hospitals and in general population samples the sensitivity of these tests for the diagnosis of alcoholism is much lower, and they identify only 10–20% of persons with less severe drinking problems. (Chick et al., 1981; Bernadt et al., 1982). All these markers are also non-specific for the detection of problem drinking, being abnormal in many disease states. Recently discovered markers, such as desialotransferrin (Stibler, Borg & Joustra, 1986; Storey et al., 1987) and antibodies to acetaldehyde-altered plasma proteins (Israel et al., 1986), may prove to be more sensitive than existing tests. Another approach is to assay alcohol in body fluids, such as saliva and sweat. This is a highly specific method, and an ‘alcohol dipstick’ has been developed for the purposes of mass screening (Kapur & Israel, 1984). The dipstick provides a rapid, semi-quantitative estimate of alcohol concentration in saliva. However, all tests based on detecting alcohol are limited to identifying consumption within the previous 24 hours and cannot in themselves estimate the severity of an underlying clinical syndrome.

At the time the present collaborative study was being devised (in 1983) there seemed to be several major limitations to the screening instrument currently available. Screening questionnaires, clinical examination procedures and laboratory tests had been developed mainly to distinguish between hospitalised alcoholics and normal drinkers. Their suitability for identifying the broad spectrum of problem drinkers was limited or had not been tested. Several appeared to be highly culturally specific, and as Room has commented, ‘off the shelf’ screening tests originating in North America, the United Kingdom and Europe may not be relevant for other cultures (Room, 1988). Thirdly, few had been developed for primary health care settings. The aim of the study was therefore to develop an instrument that would screen effectively for a broad spectrum of problem drinking and specifically for hazardous and harmful alcohol consumption. Thus, it should identify subjects before dependence, physical disease or major life problems had developed. It should be valid in different cultures. It should be suitable for primary health care, and therefore would need to be brief and easily understood to encourage its widespread use.

**Method**

**Sampling frame**

As described in the preceding paper (Saunders et al., 1993), the collaborative project involved research centres in Australia, Bulgaria, Kenya, Mexico, Norway and the USA. Subjects were recruited from representative primary health care facilities and classified on the basis of their responses in a structured interview as either ‘non drinkers’ (total abstainers or those who drank alcohol on three or fewer occasions per year and had never been treated for an alcohol problem), ‘drinking patients’ (who drank on at least four occasions per year and had never received treatment for a drinking problem) and ‘alcoholics’ (who had been diagnosed as alcoholic, had received treatment for an alcohol-related disorder or were seeking treatment). In all, 1905 subjects were interviewed in the six countries. After 17 subjects were excluded because of incomplete data on alcohol consumption, 1888 remained to form the study sample. Of these 678 (36%) were classified as non-drinkers, 913 (48%) were classified as drinking patients, and 297 (16%) were termed alcoholics. Only data obtained from the drinking patients were used to select items for AUDIT. The non-drinkers and alcoholics formed reference groups for instrument validation.

**Assessment**

Each subject was interviewed by a trained interviewer who administered a 150-item schedule which encompassed socio-demographic variables, presenting conditions, current symptomatology, past medical history, alcohol consumption, other substance use, diet, drinking behaviour, psychological reactions to alcohol, alcohol-related problems, family history of alcoholism, and self-perception of an alcohol problem (Saunders & Aasland, 1987). The interview was supplemented by a clinical examination, and blood samples were taken for
biological markers of alcohol consumption. Scores for several scales within the questionnaire were calculated according to a frequency rating (from 0 = never to 4 = daily) or a bimodal response (0 = no, 1 = yes) for individual items.

**Data analysis**
The analyses proceeded in a stepwise fashion. Firstly, the frequency distribution of responses by the drinking patients to all questions in the interview schedule was examined. Questions to which fewer than 2% of respondents answered in the affirmative were excluded from consideration for the screening instrument. Following this the intrascale reliability of each conceptual domain was determined for each national sample by computing Cronbach's alpha coefficient. A value of 0.7 or more was taken as indicating satisfactory reliability. The score for each scale was correlated with various measures of alcohol intake, including mean daily alcohol intake and frequency of having six drinks or more per session, both of which were transformed before correlation analysis by taking the natural logarithm. A correlation coefficient of 0.40 or more was taken as indicating a sufficiently close relationship with alcohol consumption to warrant further examination. For scales which fulfilled these criteria in most national samples, the item to total-minus-item correlation coefficients were calculated to identify the most representative items. A correlation matrix was then constructed of all scales and individual variables, and factor analysis was performed to examine their interrelationships and identify any domains strongly associated with alcohol intake which had not been covered by candidate questions. These analyses were carried out for each national sample and then for the aggregated data. The test-retest reliability of potential questionnaire items was then assessed, together with their validity in comparison with information from collateral informants.

Questions were not selected for AUDIT solely on the basis of statistical parameters. We made a judgement on the suitability of questions for screening and on their usefulness as a focus for therapy. Their face validity was an important consideration. Would they be understood readily by patients and provide a basis for discussion of the person's drinking problem? In addition, the value of including the major conceptual domains (alcohol consumption, drinking behaviour (dependence), and alcohol-related problems) and the adequacy of coverage of conceptual elements within a domain were taken into consideration.

**Reference standards and instrument validation**
To calculate the interim sensitivity and specificity of AUDIT, several reference standards were devised. They include hazardous alcohol consumption, which was defined for this analysis as an average daily alcohol intake exceeding 60 g per day for men and 40 g per day for women. Recurrent intoxication was defined as consumption of 60 g in a single session daily or almost daily, or 120 g per session at least weekly. 'Abnormal drinking behaviour' was diagnosed when subjects fulfilled at least one criterion of the alcohol dependence syndrome (at a minimum frequency of monthly) on the drinking behaviour scale. An alcohol dependence syndrome was diagnosed when at least three criteria were fulfilled (World Health Organization, 1992). 'Alcohol-related problems' encompassed domestic, legal and occupational difficulties related to alcohol, traumatic injury caused by drinking, and concern expressed by family, friends or health professionals. Finally, a composite index of hazardous or harmful alcohol use was constructed. A positive 'case' was defined as having any one of the following: a hazardous daily level of consumption; recurrent intoxication; abnormal drinking behaviour; at least one alcohol-related problem in the last year; an alcohol-related disease; or a perceived drinking problem. The instrument finally selected was scored from 0 to 40. Cut-off values were examined by receiver operating characteristic (ROC) analysis to identify the point(s) of maximal sensitivity and specificity with respect to hazardous and harmful alcohol use across the six samples. The sensitivity of the final instrument for detecting known alcoholics (an external reference group) was also determined.

**Results**

**Intrascale reliability of conceptual domains**
Examination of the alpha coefficients showed that among the drinking patients the drinking behaviour and adverse psychological reactions domains had high intrascale reliability, with mean values of 0.93 and 0.81 respectively.
Table 1. Item correlations and intrascale reliability coefficients for the drinking behaviour (alcohol dependence) scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Australia</th>
<th>Bulgaria</th>
<th>Kenya</th>
<th>Mexico</th>
<th>Norway</th>
<th>USA</th>
<th>Weighted mean correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Not able to stop drinking once started</td>
<td>0.49</td>
<td>0.71</td>
<td>0.92</td>
<td>0.91</td>
<td>0.87</td>
<td>0.76</td>
<td>0.81</td>
</tr>
<tr>
<td>2. Difficult to stop before became completely intoxicated</td>
<td>0.38</td>
<td>0.71</td>
<td>0.89</td>
<td>0.94</td>
<td>0.88</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>3. Fail to do what was expected because of drinking</td>
<td>0.56</td>
<td>0.64</td>
<td>0.85</td>
<td>0.93</td>
<td>0.89</td>
<td>0.53</td>
<td>0.78</td>
</tr>
<tr>
<td>4. Drunk for days</td>
<td>0.44</td>
<td>0.80</td>
<td>0.85</td>
<td>0.96</td>
<td>0.83</td>
<td>0.37</td>
<td>0.78</td>
</tr>
<tr>
<td>5. Morning drinking</td>
<td>0.49</td>
<td>0.78</td>
<td>0.64</td>
<td>0.92</td>
<td>0.79</td>
<td>0.47</td>
<td>0.73</td>
</tr>
<tr>
<td>6. Difficult to get the thought of alcohol out of your mind</td>
<td>0.40</td>
<td>0.58</td>
<td>0.82</td>
<td>0.83</td>
<td>0.84</td>
<td>0.65</td>
<td>0.71</td>
</tr>
<tr>
<td>7. Skipped meals because of drinking</td>
<td>0.55</td>
<td>0.56</td>
<td>0.81</td>
<td>0.91</td>
<td>0.69</td>
<td>0.55</td>
<td>0.70</td>
</tr>
<tr>
<td>8. Drinking at unusual times</td>
<td>0.38</td>
<td>0.79</td>
<td>0.83</td>
<td>0.90</td>
<td>0.64</td>
<td>0.45</td>
<td>0.70</td>
</tr>
<tr>
<td>9. Morning shakes</td>
<td>0.57</td>
<td>0.62</td>
<td>0.76</td>
<td>0.86</td>
<td>0.70</td>
<td>0.59</td>
<td>0.69</td>
</tr>
<tr>
<td>10. Tried to reduce consumption but failed</td>
<td>0.26</td>
<td>0.48</td>
<td>0.81</td>
<td>0.94</td>
<td>0.69</td>
<td>0.51</td>
<td>0.67</td>
</tr>
<tr>
<td>11. Needed more alcohol than before to get the desired effect</td>
<td>0.21</td>
<td>0.77</td>
<td>0.88</td>
<td>0.87</td>
<td>0.49</td>
<td>0.57</td>
<td>0.66</td>
</tr>
<tr>
<td>12. Drank more than friends</td>
<td>0.27</td>
<td>0.61</td>
<td>0.88</td>
<td>0.83</td>
<td>0.58</td>
<td>0.61</td>
<td>0.64</td>
</tr>
<tr>
<td>13. Gulping drinks to speed the effect</td>
<td>0.28</td>
<td>0.39</td>
<td>0.74</td>
<td>0.56</td>
<td>0.66</td>
<td>0.54</td>
<td>0.53</td>
</tr>
<tr>
<td>Cronbach's alpha</td>
<td>0.80</td>
<td>0.90</td>
<td>0.97</td>
<td>0.98</td>
<td>0.94</td>
<td>0.88</td>
<td>0.93</td>
</tr>
</tbody>
</table>

(Tables 1 and 2). There was relatively little variation from country to country. The values for the two domains ‘alcohol problems in the last year’ and ‘alcohol problems ever’ were lower at 0.69 and 0.65 respectively (Table 3), and varied significantly among the six samples (from approximately 0.35 to 0.83). There was a moderately strong correlation between the drinking behaviour (dependence) and mean daily alcohol consumption ($r = 0.53$) and between adverse psychological reactions, alcohol problems in the previous year and alcohol problems ever scales, and intake ($r = 0.50, 0.50$ and 0.51 respectively). The non-alcohol-specific domains such as current symptoms, past medical history and clinical examination findings had much lower values for intrascale reliability, and weaker and often non-significant correlations with mean daily alcohol intake (Saunders & Aasland, 1987). Factor analysis reported in more detail in the previous paper (Saunders et al., 1993) showed that the alcohol-specific scales and measures of consumption loaded on a dominant first factor. Current symptoms was the only non-alcohol specific domain to have a factor coefficient exceeding 0.40 (actually 0.46), and was also located in the third factor extracted, together with the clinical examination scale. On the basis of these analyses and to ensure adequate coverage of the most relevant domains, it was decided that the screening instrument should include questions on the following: (1) alcohol consumption; (2) drinking behaviour (dependence); (3) adverse psychological reactions; and (4) alcohol-related problems. It was decided not to include any questions from the non-specific domains.

**Selection of items**

From the drinking behaviour scale, three questions were selected: (1) How often during the last year have you found that you were not able to stop drinking once you had started? (2) How often . . . have you failed to do what was normally expected from you because of drinking? and (3) How often . . . have you needed a first drink in the morning to get yourself going after a heavy drinking session? The first was selected because it had the highest weighted mean item-to-total correlation coefficient (0.81), and showed the most uniform pattern of correlation across the centres (Table 1). It thus appeared to be the most representative item for the scale. It reflects impaired control over consumption of alcohol. The second question had the third highest overall correlation coefficient (0.78), and a narrow range of values from centre to centre. It represents salience of drinking, i.e. neglect of
John B. Saunders et al.

Table 2. Item correlations and intrascale reliability coefficients for the scale ‘adverse psychological reactions’

<table>
<thead>
<tr>
<th>Item Description</th>
<th>Australia</th>
<th>Bulgaria</th>
<th>Kenya</th>
<th>Mexico</th>
<th>Norway</th>
<th>USA</th>
<th>Weighted mean correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Had a feeling of remorse after drinking</td>
<td>0.38</td>
<td>0.31</td>
<td>0.73</td>
<td>0.69</td>
<td>0.56</td>
<td>0.55</td>
<td>0.53</td>
</tr>
<tr>
<td>Become depressed after drinking</td>
<td>0.49</td>
<td>0.15</td>
<td>0.70</td>
<td>0.46</td>
<td>0.21</td>
<td>0.31</td>
<td>0.44</td>
</tr>
<tr>
<td>Been unable to remember what happened the night before because you had been drinking</td>
<td>0.48</td>
<td>0.20</td>
<td>0.59</td>
<td>0.51</td>
<td>0.47</td>
<td>0.29</td>
<td>0.44</td>
</tr>
<tr>
<td>Become angry after drinking</td>
<td>0.31</td>
<td>0.26</td>
<td>0.63</td>
<td>0.46</td>
<td>0.37</td>
<td>0.59</td>
<td>0.43</td>
</tr>
<tr>
<td>Cronbach’s alpha</td>
<td>0.64</td>
<td>0.73</td>
<td>0.76</td>
<td>0.90</td>
<td>0.79</td>
<td>0.75</td>
<td>0.81</td>
</tr>
</tbody>
</table>

alternative interests and obligations in favour of drinking. The third question had the fifth highest correlation (0.73) and refers to relief of malaise after a drinking session by further consumption of alcohol. Although this may be a relatively innocuous experience at first, reflecting relief of hangover symptoms, when it occurs on a daily basis it indicates physical dependence. It was considered advantageous for both behavioural and physiological aspects of drinking behaviour to be represented. The response categories, ‘never’, ‘less than monthly’, ‘monthly’, ‘weekly’ and ‘daily’ or almost daily, are identical to those in the original assessment instrument. Responses are scored from 0 to 4 respectively.

Two questions were selected to represent adverse psychological reactions: (1) How often during the last year have you had a feeling of guilt or remorse after drinking? and (2) How often during the last year have you been unable to remember what happened the night before because you had been drinking? The first had the highest mean item-to-total coefficient in this scale (Table 2). The second showed the most uniform pattern of correlations of the three remaining questions. Reflecting as it does the experience of alcohol-induced amnesic episodes (‘blackouts’), it was considered to be a particularly useful cue for further enquiry about adverse effects of alcohol. Responses are scored as for the questions on drinking behaviour.

The two sections on alcohol-related problems (‘in the last year’ and ‘ever’) were more heterogeneous than the other alcohol domains, with mean alpha coefficients of only 0.69 and 0.65 (Table 3). The overall item-to-total correlations of the five questions were very similar (Table 3) but there was appreciable variation from country to country. Pragmatic considerations entered into the selection of items to a greater extent than elsewhere. It was decided to include a question on concern shown by family or health workers and one on alcohol related-injuries, as follows: (1) Has a relative or friend, or a doctor or other health worker, been concerned about your drinking or suggested you cut down? and (2) Have you or someone else been injured as a result of your drinking? Response categories were selected so that two different time frames were covered. An answer which indicates that the event occurred in the last year scores 4, whereas if it had happened before that, the score is 2.

The remaining questions measure alcohol consumption: (1) How often do you have a drink containing alcohol? (2) How many drinks containing alcohol do you have on a typical day when you are drinking? and (3) How often do you have six or more drinks on one occasion? They were selected as being the simplest way of capturing the usual frequency of drinking, the quantity consumed and the frequency of episodic heavy drinking. Episodic heavy drinking of this order would typically result in blood alcohol concentrations that would cause impairment of function. One ‘drink’ is assumed to contain approximately 10 g of alcohol; if in a particular culture, a typical drink contains substantially more or less than 10 g alcohol, these questions should be rephrased or the response categories altered.

Performance of the instrument
The 10 items described above form the Alcohol Use Disorders Identification Test (AUDIT).
Table 3. Item correlations and intrascale reliability coefficients for ‘alcohol problems ever’

<table>
<thead>
<tr>
<th>Item</th>
<th>Australia</th>
<th>Bulgaria</th>
<th>Kenya</th>
<th>Mexico</th>
<th>Norway</th>
<th>USA</th>
<th>Weighted mean correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Family or friend concerned about your drinking</td>
<td>0.38</td>
<td>0.31</td>
<td>0.73</td>
<td>0.69</td>
<td>0.56</td>
<td>0.55</td>
<td>0.53</td>
</tr>
<tr>
<td>2. Work difficulties because of drinking</td>
<td>0.49</td>
<td>0.15</td>
<td>0.70</td>
<td>0.46</td>
<td>0.21</td>
<td>0.31</td>
<td>0.44</td>
</tr>
<tr>
<td>3. Doctor or other health worker concerned about your drinking or suggested that you cut down</td>
<td>0.48</td>
<td>0.20</td>
<td>0.59</td>
<td>0.51</td>
<td>0.47</td>
<td>0.29</td>
<td>0.44</td>
</tr>
<tr>
<td>4. Legal trouble in connection with drinking</td>
<td>0.31</td>
<td>0.26</td>
<td>0.63</td>
<td>0.46</td>
<td>0.37</td>
<td>0.59</td>
<td>0.43</td>
</tr>
<tr>
<td>5. You or someone else injured as a result of your drinking</td>
<td>0.42</td>
<td>0.22</td>
<td>0.48</td>
<td>0.61</td>
<td>0.26</td>
<td>0.42</td>
<td>0.40</td>
</tr>
<tr>
<td>Cronbach's alpha</td>
<td>0.65</td>
<td>0.40</td>
<td>0.83</td>
<td>0.77</td>
<td>0.60</td>
<td>0.67</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Questions 1–3 measure alcohol consumption, 4–6 drinking behaviour, 7–8 adverse reactions and 9–10 alcohol-related problems. Each question is scored from 0 to 4, and the range of possible scores is from 0 to 40. The performance of AUDIT was examined in two ways, firstly by comparing scores against the diagnoses of hazardous and harmful alcohol use in the samples of drinking patients, and secondly by calculating its sensitivity among the external reference group of known alcoholics. Cut-off points for AUDIT were determined by receiver operating characteristic (ROC) analysis. The sensitivity and specificity of scores from 0 to 40 were calculated for the diagnoses of (1) hazardous daily alcohol consumption and/or recurrent intoxication, (2) abnormal drinking behaviour, (3) alcohol related problems in the last year, and (4) the combined index of hazardous and harmful alcohol consumption. Two cut-off points, 8 and 10 were identified which resulted in maximal sensitivity and specificity.

The results which are presented in Tables 4 and 5 should be regarded as interim estimates of the sensitivity and specificity of AUDIT. They represent a comparison between the AUDIT scores and diagnoses based on multiple items of information obtained during a comprehensive, structured interview, supplemented by physical examination and laboratory findings. Using the lower cut-off point of 8, the sensitivity of AUDIT for hazardous alcohol consumption ranged from 95% to 100%, for abnormal drinking behaviour from 93% to 100%, for the alcohol dependence syndrome it was 100%, and for “alcohol problems in the last year” it ranged from 91% to 100% (Table 4). The overall sensitivity for hazardous and harmful alcohol use was 87% to 96%, with an overall value of 92% (Table 4). The corresponding specificity was 81% to 98%, with an overall value of 94%. When the cut-off point of 10 was taken, sensitivity was lower (Table 5), with an overall value of 80% for the combined index. The specificity was correspondingly higher: for the combined index values ranged from 95 to 100%, and the overall value was 98%.

The validity of AUDIT was then determined among the external reference groups of known alcoholics and non-drinkers. Of the alcoholics, 99% had a score of 8 or more, 98% had a score of 10 or more and when those who were currently abstinent were excluded, all scored 10 or more. Only three of 678 non-drinkers (0.5%) had a score of 8 or more. As a final check on the accuracy of the process for selecting items, the AUDIT questions were compared with 20 other combinations of questions for their ability to discriminate between patients with hazardous or harmful alcohol use and those with non-hazardous consumption. The AUDIT questions classified subjects as accurately as any other combination.

Discussion
The development of screening instruments is necessarily linked to prevailing concepts of alcohol problems. The first screening tests were aimed at identifying alcoholics and were success-
ful in doing so. Sensitivities exceeding 95% for the MAST and its progeny attest to this. However, the ability of this family of instruments to identify patients with hazardous or harmful alcohol use is much lower (Selzer, 1971; Saunders & Kershaw, 1980; Hedlund & Vieweg, 1984). Another problem with these questionnaires is that because questions are phrased in terms of lifetime occurrence, it may be impossible to distinguish current alcoholics or problem drinkers from those who have ceased drinking or are no longer symptomatic. Indeed, this problem has led to the development of a revised response format for the MAST that provides better discrimination between current and lifetime symptoms. Reworking of the MAST has not, however, resulted in an instrument which embraces the broad spectrum of problem drinkers.

Clinical examination findings (Le Go, 1976; Skinner et al., 1986; Babor et al., 1988) have also been employed to screen for alcoholism and its physical sequelae. Although the medical impact of heavy drinking is widespread, clinical abnormalities occur relatively late in the evolution of a drinking problem (Babor, Kranzler & Lauerman, 1987) and appear to be of limited value for early intervention. Other correlates of chronic drinking, such as hypertension, are not sufficiently specific to be of much value in screening. However, they may help to confirm an impression that alcohol consumption is harmful. Established biological markers have also proved relatively insensitive in screening for hazardous alcohol consumption (Whitfield et al., 1978; Chick et al., 1981; Bernadt et al., 1982). They are also affected by substances other than alcohol as well as diseases unrelated to drinking. Serum transferrin and new immunological tests developed to measure acetaldehyde bound to plasma protein show promise as more specific markers of heavy drinking, but further research is needed to confirm their usefulness in routine screening (Stibler et al., 1986; Israel et al., 1986; Storey et al., 1987). The ideal biological marker would be one that is even more sensitive and able to identify gradations in alcohol use. Such a marker is not yet available.

Composite instruments have been devised which seek to capture a broad spectrum of physical and psychosocial consequences of alcohol use. It was hoped this would increase the sensitivity and yield of the screening process. Wilkins developed a two-stage procedure which incorporated a check-list of clinical indicators and a disguised questionnaire, the "Spare Time Activities Questionnaire" (Wilkins, 1974). The Munich Alcoholism Test (MALT) (Feuerlein et al., 1986) and the Alcohol Clinical Index (Skinner et al., 1986) also include clinical findings as well as questionnaire items. Although they are conceptually appealing and, for example, the Alcohol Clinical Index provides good discrimination between non-dependent problem drinkers and non-problem drinkers, they are relatively complex and they have not been taken up widely in clinical practice. Any method based on clinical examination findings would normally require the involvement of a medical practitioner. Likewise, most biological tests require sophisticated laboratory equipment and computer technology which might preclude their use in primary health care settings, particularly in developing countries.

Selection of items for AUDIT was guided both by the statistical analysis and certain operational requirements. The latter included the desirability of representing the conceptual domains of alcohol consumption, dependence and

<table>
<thead>
<tr>
<th></th>
<th>Hazardous consumption/reCURRENT Intoxication</th>
<th>Abnormal drinking behaviour (at least one element of dependence at specified minimum frequency)</th>
<th>Alcohol-related problems in the last year</th>
<th>Combined index of hazardous and harmful alcohol use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>100/74</td>
<td>.97/75</td>
<td>95/81</td>
<td>93/82</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>95/76</td>
<td>100/74</td>
<td>100/72</td>
<td>96/87</td>
</tr>
<tr>
<td>Kenya</td>
<td>97/66</td>
<td>.97/69</td>
<td>100/71</td>
<td>95/81</td>
</tr>
<tr>
<td>Mexico</td>
<td>100/77</td>
<td>93/89</td>
<td>94/98</td>
<td>88/98</td>
</tr>
<tr>
<td>Norway</td>
<td>100/89</td>
<td>94/91</td>
<td>91/94</td>
<td>87/97</td>
</tr>
<tr>
<td>USA</td>
<td>96/74</td>
<td>.97/79</td>
<td>92/83</td>
<td>90/92</td>
</tr>
<tr>
<td>All countries</td>
<td>97/78</td>
<td>.96/81</td>
<td>95/85</td>
<td>92/94</td>
</tr>
</tbody>
</table>
Table 5. Sensitivity and specificity of the ‘AUDIT’ questionnaire at cut-off point of 10

<table>
<thead>
<tr>
<th>Hazardous consumption/recurrent intoxication</th>
<th>Abnormal drinking behaviour (at least one element of dependence at specified minimum frequency)</th>
<th>Alcohol-related problems in the last year</th>
<th>Combined index of hazardous and harmful alcohol use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>97/86</td>
<td>85/92</td>
<td>79/98</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>84/87</td>
<td>84/82</td>
<td>79/96</td>
</tr>
<tr>
<td>Kenya</td>
<td>94/79</td>
<td>97/84</td>
<td>91/95</td>
</tr>
<tr>
<td>Mexico</td>
<td>100/83</td>
<td>84/100</td>
<td>79/100</td>
</tr>
<tr>
<td>Norway</td>
<td>90/94</td>
<td>74/98</td>
<td>67/99</td>
</tr>
<tr>
<td>USA</td>
<td>86.81</td>
<td>87/91</td>
<td>76/96</td>
</tr>
<tr>
<td>All countries</td>
<td>92/87</td>
<td>86/90</td>
<td>80/98</td>
</tr>
</tbody>
</table>

problems, the need for a rounded number of questions and for all questions to be easily understandable, valid across different cultures, and capable of providing the framework for subsequent intervention. Analysis revealed that in some scales (for example, drinking behaviour) many items were virtually interchangeable in terms of their correlation with total scores, or capacity to distinguish between hazardous and harmful drinkers and those in the non-hazardous range. The final 10-item instrument included three questions on intake, four on alcohol-related problems and adverse reactions, and three on drinking behaviour. Because these domains conform approximately to the concepts of hazardous alcohol use, harmful use and dependence in WHO terminology and the ICD-10 psychoactive substance use disorders section (World Health Organization, 1992), the test was named the Alcohol Use Disorders Identification Test (AUDIT).

The differences between AUDIT and most existing questionnaires can be summarised as follows. Firstly, it seeks to detect problem drinkers at the less severe end of the spectrum, rather than targeting persons with established dependence or alcoholism. Secondly, it places considerable emphasis on hazardous consumption and frequency of intoxication compared with drinking behaviour and adverse consequences. It refers to alcohol experiences in the past year as well as lifetime experience. This improves its relevance to current drinking status. It does not require the individual to identify as a problem drinker. The responses are not yes/no ones but are based on the frequency of the experience, and range from ‘never’ and ‘less than monthly’ through to daily. It is anticipated that this will reduce under-reporting of adverse effects. Not surprisingly, there are also some similarities with existing instruments. Two of the ten items (guilt about drinking, and drinking in the morning) are similar to those in the CAGE (Ewing, 1984), and five resemble questions in the MAST (Selzer, 1971). AUDIT is shorter than both the full MAST and a more recently introduced instrument with a similar conceptual basis, the Canterbury Alcoholism Screening Test (Elvy & Wells, 1984). It can be embedded within general health and lifestyle questionnaires without making them unwieldy or upsetting their balance. This would provide an element of disguise which may be useful in some settings.

The comparative merit for screening purposes of “consumption” questions and “adverse consequences” ones has been the subject of debate. Ryder and colleagues reported that questions on alcohol-related problems were more sensitive indicators of problem drinking than consumption ones (Ryder et al., 1988). In an antenatal population, however, quantity-frequency questions gave a higher yield than the CAGE or Brief MAST (Waterson & Murray-Lyon, 1989). The main conclusion from these and related studies is that a screening instrument for general purposes should combine consumption and consequences questions (Barrison et al., 1982; Cutler, Wallace & Haines, 1988; Cyr & Wartman, 1988; Persson & Magnusson, 1988; Babor et al., 1989a).

The unique feature of AUDIT is that it has been derived from a cross-national data set. Only those questions which could be translated, literally and idiomatically, into multiple languages were included in the original assessment schedule. Questions were selected on their repre-
sentativeness in the pooled data set and checks were made to ensure that none performed poorly in an individual national sample. The sensitivity and specificity of AUDIT (Tables 4 and 5) are similar from country to country, and there is no evidence of dominance by one particular culture as judged by these parameters. Experience using existing screening instruments in other cultures has often been unsatisfactory. Only five of the 31 criteria in the Munich Alcoholism Test, derived in Germany, were found to be 'relatively free of cultural differences' when the test was applied in Spain and Ecuador (Gorenc et al., 1984). Use of an alcoholism questionnaire in an Alaskan Eskimo population resulted in 75% of the adult population being classified as 'probable alcoholics', a proportion considered to be far in excess of actuality (Klausner & Foulks, 1982). The problem is less when there is careful selection of items and adequate checks on translation: the MAST discriminated well between alcoholic and non-alcoholic subjects in an Italian sample (Garzotto et al., 1988). One can be optimistic that AUDIT will prove useful in countries with similar cultural, political and economic characteristics to those represented in the present study.

The validity of AUDIT was examined in relation to its ability to discriminate between persons with harmful or hazardous alcohol consumption and those with non-hazardous consumption. This approach brings certain problems in its wake. Definitions of hazardous alcohol intake, recurrent intoxication, abnormal drinking behaviour and alcohol-related problems had to be established by the collaborating investigators since there were no internationally agreed criteria at the time the analyses were undertaken. They were based on the level of consumption or harm above which intervention was judged to be preferable to no intervention. A composite reference standard was also defined which encompassed these entities and also included the presence of an alcohol-related disease and acknowledgement by the subject of a drinking problem. This reference standard was judged to incorporate the key elements of hazardous and harmful consumption, as defined by WHO and in the ICD-10 system, then in the process of development.

Calculations of sensitivity and specificity in the present report should be supplemented by results of further field studies. The questions which form AUDIT were embedded within the assessment schedule and were not administered separately. An individual's response to questions which are grouped together may differ from when they form part of a larger schedule. In practice, however, AUDIT may be used more frequently when embedded within a broadly-based lifestyle questionnaire than as a discrete alcohol screening instrument. Secondly, a composite question was formed to inquire of concern shown by family or health professionals. Thirdly, the response categories for the two questions on alcohol-related problems differ from those used in the assessment instrument. Finally, the criteria against which AUDIT was judged (e.g. hazardous consumption, abnormal drinking behaviour, alcohol-related problems) are represented in it. Validating a new screening test by reference to a broader assessment schedule which incorporates the same key elements may result in classification accuracy and validity coefficients being inflated (Babor & Kadden, 1985). Nevertheless, the present study included an external criterion group (known alcoholics) and AUDIT performed well with this sample.

It is now appropriate to submit the instrument to extensive field testing. A user's manual and research agenda have been developed for this purpose (Babor et al., 1989b). Among the questions that need to be answered are its accuracy and usefulness in different health care facilities, such as general practitioners' (family physicians') rooms, community health clinics, hospital outpatient clinics, and inpatient services. It may also find a place in epidemiological surveys of general (i.e. non-clinical) populations. Although the countries represented in the present project were culturally diverse, it would be appropriate in the next phase of work to include additional cultures. The issue of whether AUDIT should be administered as it exists at present or incorporated into a larger questionnaire which enquires of other lifestyle issues (such as diet, cigarette smoking and exercise), merits exploration: a degree of disguise to the purpose of screening may facilitate accurate self-reporting. Ultimately, the predictive validity of the instrument needs to be documented. Reassessment of sub-samples of subjects in the present study is presently being undertaken three years after they were originally interviewed.

At the outset of the study it was hoped to identify questions which did not mention alcohol
explicitly which would prove to be useful for identifying harmful alcohol consumption. It was considered that such questions would be unaffected by response bias in self-reports on alcohol use and consequences. This did not eventuate. Correlations between gastrointestinal and other physical symptoms and psychological disturbances, and measures of alcohol intake were generally low or non-significant. There did not appear to be adequate grounds for including them in a screening instrument.

The present study is part of a WHO collaborative programme on early intervention. The ultimate aim is to make available a valid screening procedure, and an effective and robust form of brief therapy which can be applied at the point of first contact. This approach capitalises on the accessibility of primary health care and the high throughput of patients, and eliminates the problem of attrition where referral to a specialist service is the only option. More fundamentally, it identifies the primary health care worker as the key person in the strategy to reduce alcohol-related harm throughout the community.

Acknowledgements
This paper is based on the data and experience obtained during the participation of the authors in the WHO Collaborative Projects on Detection and Treatment of Persons with Harmful Alcohol Consumption, sponsored by the World Health Organization and the participating field research centres. The collaborating investigators in this study are: at Headquarters in WHO, Geneva: Ms K. Canavan, Mr M. Grant, Mr J. Ordin and Dr N. Sartorius; in the field research centres in Australia: Dr F. Harding Burns, Dr R. Reznik and Dr J. B. Saunders; in Bulgaria: Dr M. Boyadjieva; in Kenya: Dr S. W. Acuda; in Mexico: Dr J. R. de la Fuente; in Norway: Dr O. Aasland, Mr A. Amundsen and Dr J. Morland; and in the USA: Dr T. Babor, Dr H. Kranzler and Dr R. Lauerman. The views expressed in this paper are the authors' own and do not necessarily represent the opinions of the other investigators participating in the project nor the views or policies of sponsoring agencies. The advice and comments of Dr P. Anderson (Intstat Australia) are much appreciated.

References


**Institute of Medicine** (1990) Broadening the Base of Treatment for Alcohol Problems (Washington, DC, National Academy Press).


Appendix

AUDIT

Please circle the answer that is correct for you

1. How often do you have a drink containing alcohol?
   - Never
   - Monthly
   - Two to four times a month
   - Two to three times a week
   - Four or more times a week

2. How many drinks containing alcohol do you have on a typical day when you are drinking?
   - 1 or 2
   - 3 or 4
   - 5 or 6
   - 7 to 9
   - 10 or more

3. How often do you have six or more drinks on one occasion?
   - Never
   - Less than monthly
   - Monthly
   - Weekly
   - Daily or almost daily

4. How often during the last year have you found that you were not able to stop drinking once you had started?
   - Never
   - Less than monthly
   - Monthly
   - Weekly
   - Daily or almost daily

5. How often during the last year have you failed to do what was normally expected from you because of drinking?
   - Never
   - Less than monthly
   - Monthly
   - Weekly
   - Daily or almost daily

6. How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?
   - Never
   - Less than monthly
   - Monthly
   - Weekly
   - Daily or almost daily

7. How often during the last year have you had a feeling of guilt or remorse after drinking?
   - Never
   - Less than monthly
   - Monthly
   - Weekly
   - Daily or almost daily

8. How often during the last year have you been unable to remember what happened the night before because you had been drinking?
   - Never
   - Less than monthly
   - Monthly
   - Weekly
   - Daily or almost daily

9. Have you or someone else been injured as a result of your drinking?
   - No
   - Yes, but not in the last year
   - Yes, during the last year

10. Has a relative or friend, or a doctor or other health worker been concerned about your drinking or suggested you cut down?
    - No
    - Yes, but not in the last year
    - Yes, during the last year
Procedure for Scoring AUDIT

Questions 1–8 are scored 0, 1, 2, 3 or 4. Questions 9 and 10 are scored 0, 2 or 4 only. The response coding is as follows:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>Never</td>
<td>Monthly or less</td>
<td>Two to four times per month</td>
<td>Two to three times per week</td>
<td>Four or more times per week</td>
</tr>
<tr>
<td>Question 2</td>
<td>1 or 2</td>
<td>3 or 4</td>
<td>5 or 6</td>
<td>7 to 9</td>
<td>10 or more</td>
</tr>
<tr>
<td>Questions 3–8</td>
<td>Never</td>
<td>Less than monthly</td>
<td>Monthly</td>
<td>Weekly</td>
<td>Daily or almost daily</td>
</tr>
<tr>
<td>Questions 9–10</td>
<td>No</td>
<td>Yes, but not in the last year</td>
<td>Yes, during the last year</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The minimum score (for non-drinkers) is 0 and the maximum possible score is 40.
A score of 8 or more indicates a strong likelihood of hazardous or harmful alcohol consumption.
This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.