

SCREENING FOR ALCOHOL USE DISORDERS AND AT-RISK DRINKING IN THE GENERAL POPULATION: PSYCHOMETRIC PERFORMANCE OF THREE QUESTIONNAIRES

HANS-JÜRGEN RUMPF*, ULFERT HAPKE¹, CHRISTIAN MEYER¹ and ULRICH JOHN¹

Medical University of Lübeck, Department of Psychiatry and Psychotherapy, Research Group S:TEP (Substance Abuse: Treatment, Epidemiology, and Prevention), Lübeck and ¹University of Greifswald, Institute of Epidemiology and Social Medicine, Addiction Research Center, Greifswald, Germany

(Received 29 December 2000; in revised form 7 September 2001; accepted 18 November 2001)

Abstract — Most screening questionnaires are developed in clinical settings and there are few data on their performance in the general population. This study provides data on the area under the receiver-operating characteristic (ROC) curve, sensitivity, specificity, and internal consistency of the Alcohol Use Disorders Identification Test (AUDIT), the consumption questions of the AUDIT (AUDIT-C) and the Lübeck Alcohol Dependence and Abuse Screening Test (LAST) among current drinkers ($n = 3551$) of a general population sample in northern Germany. Alcohol dependence and misuse according to DSM-IV and at-risk drinking served as gold standards to assess sensitivity and specificity and were assessed with the Munich-Composite Diagnostic Interview (M-CIDI). AUDIT and LAST showed insufficient sensitivity for at-risk drinking and alcohol misuse using standard cut-off scores, but satisfactory detection rates for alcohol dependence. The AUDIT-C showed low specificity in all criterion groups with standard cut-off. Adjusted cut-points are recommended. Among a subsample of individuals with previous general hospital admission in the last year, all questionnaires showed higher internal consistency suggesting lower reliability in non-clinical samples. In logistic regression analyses, having had a hospital admission increased the sensitivity in detecting any criterion group of the LAST, and the number of recent general practice visits increased the sensitivity of the AUDIT in detecting alcohol misuse. Women showed lower scores and larger areas under the ROC curves. It is concluded that setting specific instruments (e.g. primary care or general population) or adjusted cut-offs should be used.

INTRODUCTION

Several screening questionnaires have been developed to detect individuals with alcohol dependence, alcohol misuse or high levels of alcohol consumption. Mostly, these screening tools are applied in medical care settings to detect subjects eligible for brief interventions. In general population studies, screening instruments serve as tools for case finding or to provide data to estimate prevalence rates of alcohol problems, and are used when other diagnostic procedures are too comprehensive. Screening questionnaires in general population studies are appealing, because of their inexpensive format. However, it might be argued that screening questionnaires developed in clinical settings are not automatically suitable in the general population and that a spectrum bias may occur, e.g. because of differences in the severity of dependence.

Only a few studies have addressed the validity of screening questionnaires in the general population. The CAGE (Ewing, 1984; Mayfield *et al.*, 1974) (acronym based on its four items: Cut down on drinking, Annoyed by criticism, Guilty feelings, and Eye opener) revealed a lower sensitivity in the general population than in primary-care patients (Chan *et al.*, 1994a; Cherpitel, 1998) and in an emergency room setting (Cherpitel, 1998). Interestingly, among the general population sample of the latter study, the CAGE showed a tendency to perform better among individuals reporting an emergency room or primary-care visit in the previous 12 months and was significantly more sensitive in men with a previous emergency room visit in the last year (Cherpitel, 1999). The lower validity of the CAGE in general population samples corresponds with findings from a large scale Canadian study (Bisson *et al.*, 1999). The Brief MAST (Pokorny *et al.*, 1972), a shortened version of the

Michigan Alcoholism Screening Test (MAST; Selzer, 1971), was also less sensitive in a general population sample compared with primary-care out-patients (Chan *et al.*, 1994b). For the TWEAK test (Russell *et al.*, 1994) (acronym based on its five items Tolerance, Worry about drinking, Eye opener, Amnesia (blackouts), and c(K)ut down on drinking), findings are not so clear. In one study, the sensitivity of the TWEAK was lower in the general population, compared to an emergency room sample, but higher compared to primary-care patients, in identifying alcohol dependence (Cherpitel, 1998). In a second study, no differences in sensitivity were found between a general population and a primary-care sample in detecting heavy drinking (Chan *et al.*, 1993). Using alcohol dependence as gold standard, differences in sensitivity of the TWEAK between samples depended on two versions of the tolerance item. In summary, there is evidence that screening questionnaires show different psychometric properties in the general population, compared to samples drawn in medical settings. No data with respect to the validity in the general population could be found for three more recently developed instruments: the Alcohol Use Disorders Identification Test (AUDIT; Babor *et al.*, 1989b; Saunders *et al.*, 1993), the AUDIT Alcohol Consumption Questions (AUDIT-C; Bush *et al.*, 1998) and the Lübeck Alcohol Dependence and Abuse Screening Test (LAST; Rumpf *et al.*, 1997). The AUDIT has been used in the general population; however, data are restricted to subgroups (unemployed; Claussen and Aasland, 1993) and do not give clear estimates of validity such as sensitivity and specificity based on a gold standard (Holmila, 1995; Fleming, 1996; Medina-Mora *et al.*, 1998).

The aims of the present study were: (1) to assess and compare the performance of the AUDIT, the AUDIT-C and the LAST in a general population sample; (2) to examine different cut-off points for the three instruments; (3) to analyse age and gender effects; (4) to test whether sensitivity and internal consistency varied in the subsamples of individuals reporting general hospital admissions or general practice visits in the previous 12 months.

*Author to whom correspondence should be addressed at: Department of Psychiatry and Psychotherapy, Research Group S:TEP, Medical University of Ratzeburger Allee 160, 23538 Lübeck, Germany.

SUBJECTS AND METHODS

Population area of study

The study was part of the project on Transitions in Alcohol Consumption and Smoking (TACOS; Hapke *et al.*, 1998; Rumpf *et al.*, 1998a). The present data are derived from a general population sample in Lübeck, a northern German city with 217 000 inhabitants, and 46 adjoining communities. Individuals born between 1932 and 1978 were randomly drawn from the official resident registration office files (representing the age group 18–64 years in the mid of data gathering). In Germany, residents are bound by law to register within 4 weeks after moving to a new place. Therefore, these files are a valuable source to obtain representative samples. All individuals with German nationality (to avoid language problems) and not living in institutions were included in the study. The study area is ~30 miles across with a population of 325 000 individuals (including the city of Lübeck and the 46 adjoining communities which represent the catchment area of Lübeck). Two-thirds of the sample were conducted in Lübeck and one-third in the adjoining communities. Data were collected during 8 months (July 1996 to March 1997) by computer-assisted personal interviews. Interviews were conducted by 56 trained lay-interviewers.

Sample

From 6447 addresses drawn from the office files, 618 (9.6%) were invalid for various reasons (e.g. the individual moved away, was not known in the household, deceased, or did not fulfil the inclusion criteria). Of the remaining 5829 valid addresses, 665 (11.4%) individuals were not available during the study period, 83 (1.4%) did not participate because of being ill, 979 (16.8%) refused to participate, and nine (0.2%) refused an interview other than by telephone or only partially completed the interview. This results in a response rate of 70.2%. Of these 4093 interviews, 18 could not be analysed mainly because of technical problems. Therefore, the sample consisted of 4075 respondents. Only individuals consuming alcohol within the last 12 months ($n = 3641$) were asked to fill out the AUDIT and the LAST. Seventy-nine individuals had at least one missing value in AUDIT or LAST and 11 had missing values with respect to the recency of the alcohol-related disorders or in quantity–frequency questions, and were thus excluded from the analysis, which resulted in a final sample of 3551 analysed subjects. Of this sample, 50.8% were male, 45% had up to 9 years of schooling, 32.8% 10–11 years, and 22.1% 12 or more years. Mean (\pm SD) age was 41.2 ± 12.8 years. Of the sample, 59.6% were married, 28.5% never married, and 11.9% widowed, divorced or separated.

Diagnosis of alcohol-related disorders

Alcohol dependence according to DSM-IV (American Psychiatric Association, 1995) was assessed by the Munich–Composite International Diagnostic Interview (M-CIDI; Wittchen *et al.*, 1995) the German version of the CIDI (Robins *et al.*, 1988). Interviewers were trained by World Health Organization–CIDI trainers in a 1-week course. Five psychologists supervised and edited the interviews. Of the sample described above, 1.4% fulfilled DSM-IV criteria in the last 12 months for current alcohol dependence and 1.2% for current alcohol misuse according to M-CIDI. For alcohol use

disorders according to M-CIDI, test–retest reliability was found to be excellent (Wittchen *et al.*, 1998).

Following criteria of the British Medical Association (1995), at-risk drinking was defined as average daily consumption of at least 20 (women) or 30 (men) g of pure alcohol. Alcohol consumption was assessed by using the quantity–frequency questions of the M-CIDI. Individuals drinking alcohol more than 12 times in the last 12 months were defined as current drinkers. Among these individuals, frequency of alcohol consumption was assessed by using the following categories: almost daily, 3–4 times per week, 1–2 times per week, 1–3 times per month, less often than monthly. Quantity assessment was supported by a visual aid showing typical alcoholic beverages. One standard drink was converted into 9 g of pure alcohol. A quantity–frequency index was computed by using the mean of the categories. The quantity–frequency assessment of the M-CIDI showed excellent test–retest reliability (Lachner *et al.*, 1998). Among the sample described above, 5.4% fulfilled criteria for at-risk drinking.

These rates of at-risk drinking, alcohol dependence, and alcohol misuse are rather small compared to data from the USA and are lower than in other regions of Germany. This is due to distinct regional differences in alcohol consumption between federal states in Germany with rates of at-risk drinking ranging between 2.2 and 23% (mean: 13.5%). Lübeck belongs to the state with the second lowest rate of at-risk drinkers (Meyer *et al.*, 1998).

Screening questionnaires

German versions of AUDIT and LAST were presented as self-administered questionnaires at the end of the interview. The English language version items and the scoring are presented in Table 1. The alcohol use disorders section of the M-CIDI was presented after the sections on sociodemographics, tobacco use, affective, anxiety, somatoform, and eating disorders. Between the alcohol section of the M-CIDI and the screening questionnaires, comprehensive questions not related to alcohol were presented including the CIDI sections on obsessive–compulsive disorder, illicit drugs, and post-traumatic stress disorder, followed by questions on health-care utilization and several self-administered questionnaires with 125 items on mental health, sense of coherence, satisfaction with life, social support, nutrition, and physical activities. In total, questions on mental health far outnumbered questions on alcohol use.

The AUDIT was developed in a World Health Organization project to provide a tool suitable for the detection of problem drinkers in primary-care settings. In contrast to other frequently used instruments, such as CAGE or MAST, the AUDIT aims to detect individuals with hazardous alcohol use or at-risk drinking, rather than alcohol dependence or alcohol misuse. The AUDIT core questionnaire consists of 10 items: three questions on quantity and frequency of drinking, three items on alcohol dependence and four questions on problems caused by drinking. Weighted scoring with respect to the frequency or the time of occurrence results in total scores ranging from 0 to 40. In the original study, a score of 8 points was recommended as a cut-off (Babor *et al.*, 1989a). Recommendations in other studies range from 5 (Schmidt *et al.*, 1995) to 10 (Bohn *et al.*, 1995). Using the usual cut-off of 8, the sensitivity (rate of correctly identified positive cases) for identifying individuals with alcohol dependence or misuse ranges between 0.38

Table 1. The Alcohol Use Disorders Identification Test (AUDIT), the Alcohol Use Disorders Identification Test Consumption Questions (AUDIT-C), and the Lübeck Alcohol Dependence and Abuse Screening Test (LAST)

Item	Scoring				
AUDIT					
1. How often do you have a drink containing alcohol?	Never (0)	Monthly or less (1)	2-4 times a month (2)	2-3 times a week (3)	≥4 times a week (4)
2. How many drinks containing alcohol do you have on a typical day when you are drinking?	1 or 2 (0)	3 or 4 (1)	5 or 6 (2)	7 to 9 (3)	≥10 (4)
3. How often do you have six or more drinks on one occasion?	Never (0)	Less than monthly (1)	Monthly (2)	Weekly (3)	Daily or almost daily (4)
4. How often during the last year have you found that you were not able to stop drinking once you have started?	Never (0)	Less than monthly (1)	Monthly (2)	Weekly (3)	Daily or almost daily (4)
5. How often during the last year have you failed to do what was normally expected from you because of drinking?	Never (0)	Less than monthly (1)	Monthly (2)	Weekly (3)	Daily or almost daily (4)
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 (0)	Less than monthly (1)	Monthly (2)	Weekly (3)	Daily or almost daily (4)
7. How often during the last year have you had a feeling of guilt or remorse after drinking?	Never (0)	Less than monthly (1)	Monthly (2)	Weekly (3)	Daily or almost daily (4)
8. How often during the last year have you been unable to remember what happened the night before because you had been drinking?	Never (0)	Less than monthly (1)	Monthly (2)	Weekly (3)	Daily or almost daily (4)
9. Have you or someone else been injured as a result of your drinking?	No (0)			Yes, but not in the last year (2)	Yes, during the last year (4)
10. Has a relative, friend, doctor, or other health worker been concerned about your drinking or suggested that you should cut down?	No (0)			Yes, but not in the last year (2)	Yes, during the last year (4)
LAST					
1. Are you always able to stop drinking when you want to?	Yes (0)	No (1)			
2. Have you ever felt you should cut down on your drinking?	Yes (1)	No (0)			
3. Have you ever felt bad or guilty about your drinking?	Yes (1)	No (0)			
4. Does your wife, husband, a parent, or other near relative ever worry or complain about your drinking?	Yes (1)	No (0)			
5. Have you ever gotten into trouble at work because of drinking?	Yes (1)	No (0)			
6. Have you ever been told you have liver trouble? Cirrhosis?	Yes (1)	No (0)			
7. Have you ever been in a hospital because of drinking?	Yes (1)	No (0)			

The AUDIT-C comprises items 1 to 3 of the AUDIT.

(Schmidt *et al.*, 1995) and 0.96 (Isaacson *et al.*, 1994) with a specificity (rate of correctly identified negative cases) of 0.95 and 0.96, respectively. Research on the AUDIT was reviewed by Allen *et al.* (1997).

The AUDIT-C (short for AUDIT consumption questions) includes the first three items of the original instrument (Bush *et al.*, 1998). Using a cut-off of 3 points, the AUDIT-C revealed a sensitivity of 0.90 for active alcohol misuse or dependence and 0.98 for heavy drinking (>14 drinks a week or ≥5 drinks on one occasion) with a rather low specificity (0.60). A higher cut-off resulted in a sensitivity of 0.86 and a specificity of 0.72 of patients with heavy drinking or alcohol dependence or misuse. The AUDIT-C outperformed the full AUDIT when identifying heavy drinking, but was inferior for alcohol misuse or dependence. Findings are restricted to male subjects and, to date, no further validation is available. In our study, the AUDIT-C scores were calculated by using the first three items of the full AUDIT.

The LAST was developed in a general hospital sample by combining the instruments CAGE and MAST (Rumpf *et al.*, 1997). This questionnaire consists of seven dichotomous items (two from the CAGE and five from the MAST) and is scored without using weightings, with 2 points as cut-off. The LAST revealed a higher sensitivity, compared to the CAGE and the 13-item Short Michigan Alcoholism Screening Test (SMAST; Selzer *et al.*, 1975) and showed no significant differences in sensitivity when compared to the more comprehensive MAST

(Rumpf *et al.*, 1997). The sensitivity in the detection of patients with alcohol dependence or misuse ranged from 0.63 (general practice) to 0.87 (general hospital). The specificity ranged between 0.88 (general hospital) and 0.93 (general practice). Compared to AUDIT and AUDIT-C, the LAST incorporates two items with more clinical aspects ('Have you ever been told you have liver trouble? Cirrhosis?', 'Have you ever been in a hospital because of drinking?').

Data analysis

The concurrent validity of the screening questionnaires was assessed by calculating sensitivity (rate of correctly identified individuals having the respective disorder) and specificity (rate of correctly identified individuals not having the respective disorder). Differences in sensitivity between tests were analysed using the non-parametric McNemar test for two related samples. Receiver-operating characteristics (ROC) curves were computed using SPSS 9.0; the area under the curve was used to compare the performance of the instruments by additional calculations (McClish, 1991). ROC curves allow the exploration of the entire range of sensitivities and specificities at each possible cut-off point by showing sensitivity at the y-axis and (1 - specificity) at the x-axis. Cut-off point decisions were made on grounds of the ROC curves. Those cut-offs were chosen where concavities occurred, otherwise the closest distance of the curve to the upper left corner was sought, and the cut-point above this was chosen.

To compare the quantity–frequency assessment of AUDIT and M-CIDI as a measure of concurrent validity, a quantity–frequency index was computed by using the mean of the categories of the AUDIT questions as multiplier (AUDIT 1: 0, 0.033, 0.1, 0.357, 0.786; AUDIT 2: 1.5, 3.5, 5.5, 8, 10) Logistic regression analyses were used to examine the impact of gender and previous health-care utilization on the sensitivity of the screeners. Cronbach's alpha was calculated to assess internal consistency.

RESULTS

Concurrent validity

Table 2 displays the areas under the ROC curves showing the performance of the screeners over the full range of its scores. Areas are highest for alcohol dependence and lowest for alcohol misuse.

Sensitivity and specificity are shown in Table 3. Using the recommended cut-off of 8, the AUDIT showed low sensitivity for at-risk drinking (0.33), current alcohol misuse (0.37), and any criterion (0.41), but a good sensitivity for current alcohol dependence (0.78), whereas specificity was high, ranging between 0.94 and 0.96 (Table 3). The AUDIT-C showed low specificity when using recommended cut-off points of 3 or 4

(Table 3). The LAST showed low sensitivity using the standard cut-off score of 2 for all diagnostic groups. Comparing the quantity–frequency index assessed by AUDIT and AUDIT-C with that of the M-CIDI showed a high correlation (Spearman rho) of 0.83 which was significant ($P < 0.001$).

Based on analyses of ROC curves (see 'Data analysis'), the following cut-offs seem appropriate in this sample: the AUDIT performs best with ≥ 5 points as a positive result for at-risk drinking, 4 points for misuse and 6 points for dependence, the AUDIT-C with 5 points for at-risk drinking and dependence, and 4 points for misuse, and the LAST with 1 point for all groups. However, cut-off points for alcohol misuse result in rather low specificities for AUDIT (0.60) and AUDIT-C (0.62); therefore, a cut-off of 5 is considered more appropriate for both questionnaires in this criterion group. Using these cut-off points leads to more satisfactory validity measures for all instruments (Table 4).

Differences in validity between tests

Comparing areas under the curve shows that over the full range of scores AUDIT was not significantly different from AUDIT-C for at-risk drinking ($\chi^2 = 0.61$; $df = 1$; $P = 0.44$), alcohol misuse ($\chi^2 = 0.37$; $df = 1$; $P = 0.54$), and alcohol dependence ($\chi^2 = 3.15$; $df = 1$; $P = 0.08$); AUDIT showed larger areas under the curve, compared to LAST for at-risk

Table 2. Performance of AUDIT, AUDIT-C and LAST for different diagnostic groups

Diagnostic group	AUDIT		AUDIT-C		LAST	
	Area under ROC curve	95% CI	Area under ROC curve	95% CI	Area under ROC curve	95% CI
At-risk drinking ($n = 191$)	0.86	0.84–0.88	0.87	0.85–0.89	0.63	0.58–0.67
Current alcohol misuse ($n = 41$)	0.81	0.76–0.87	0.79	0.73–0.84	0.67	0.57–0.76
Current alcohol dependence ($n = 49$)	0.96	0.94–0.98	0.93	0.89–0.96	0.89	0.83–0.95
Any criterion ^a ($n = 281$)	0.89	0.87–0.90	0.88	0.87–0.90	0.69	0.65–0.73

n (total sample) = 3551.

^aCurrent at-risk drinking, alcohol misuse or dependence.

ROC = receiver-operating characteristic curve; CI = confidence interval. For other abbreviations, see Table 1.

Table 3. Sensitivity and specificity of AUDIT, AUDIT-C and LAST for different cut-off scores

Cut-off	At-risk drinking ($n = 191$)		Current alcohol misuse ($n = 41$)		Current alcohol dependence ($n = 49$)		Any criterion ^a ($n = 281$)	
	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
AUDIT								
4	0.95	0.63	0.90	0.60	1.0	0.60	0.95	0.64
5	0.77	0.80	0.61	0.77	0.96	0.78	0.78	0.81
6	0.54	0.89	0.49	0.87	0.92	0.88	0.60	0.91
7	0.41	0.93	0.41	0.91	0.88	0.92	0.49	0.94
8	0.33	0.95	0.37	0.94	0.78	0.94	0.41	0.96
AUDIT-C								
3	0.99	0.42	0.95	0.40	1.0	0.40	0.99	0.43
4	0.94	0.65	0.83	0.62	0.96	0.62	0.93	0.66
5	0.74	0.83	0.56	0.81	0.88	0.81	0.74	0.85
LAST								
1	0.42	0.83	0.49	0.82	0.86	0.83	0.51	0.85
2	0.24	0.92	0.34	0.92	0.71	0.92	0.34	0.94

n (total sample) = 3551.

^aCurrent at-risk drinking, alcohol misuse or dependence.

Sens., sensitivity; Spec. (specificity); for other abbreviations, see Table 1.

drinking ($\chi^2 = 86.31$; $df = 1$; $P < 0.001$), alcohol misuse ($\chi^2 = 6.60$; $df = 1$; $P < 0.05$), and alcohol dependence ($\chi^2 = 5.30$; $df = 1$; $P < 0.05$); AUDIT-C was not different from LAST for alcohol dependence ($\chi^2 = 0.92$; $df = 1$; $P = 0.26$), but showed larger areas for at-risk drinking ($\chi^2 = 94.65$; $df = 1$; $P < 0.001$) and alcohol misuse ($\chi^2 = 4.36$; $df = 1$; $P < 0.05$).

Table 4 displays differences in sensitivity and specificity between tests at cut-off points found to be appropriate in the present sample. For this comparison, the following cut-off points were used. For the AUDIT, 5 was used as a cut-off point in all subgroups, except for alcohol dependence (cut-off: 6). For the AUDIT-C, 5 was used as the cut-off point with regard to each criterion, and for the LAST, 1 point served as the threshold for all three criteria. The AUDIT performed best for at-risk drinking, showing significantly higher sensitivity, but specificity was lower compared to AUDIT-C and LAST. For current alcohol misuse and alcohol dependence, no significant differences in sensitivity were found between tests. The LAST outperforms both AUDIT versions in specificity for alcohol misuse and the AUDIT shows better specificity for alcohol dependence, compared to the other tests. Using any of the three criteria, the AUDIT outperforms AUDIT-C and LAST in sensitivity, whereas AUDIT-C and LAST show higher specificity.

Internal consistency

Cronbach's alpha was used as a measure of internal consistency. The AUDIT showed a moderate alpha of 0.75. The lowest corrected item-total correlation was found for item 1 (frequency of alcohol consumption: 0.28) and item 9 (injured as a result of drinking: 0.36), all other items ranged between 0.48 and 0.58. The AUDIT-C, having only three items, revealed low internal consistency with an alpha of 0.56. Corrected item-total correlations ranged from 0.30 (item 1: frequency) to 0.52 (item 3: ≥ 6 drinks on one occasion). The LAST showed a moderate Cronbach's alpha of 0.72. The lowest corrected item-total correlation was found for item 1 (always able to stop drinking: 0.28) and item 6 (been told of having liver problems: 0.36), all other item-total correlations ranged between 0.44 and 0.58.

Impact of previous general hospital admissions and general practice visits on sensitivity and internal consistency

To examine if questionnaires have different sensitivities for individuals with previous primary-care contacts, logistic

regression analyses were performed separately for the three questionnaires. As dependent variable, the sensitivity of the respective test (0 = not identified; 1 = identified), separately for each of the three criterion groups, was used. As independent variable, two dichotomous variables were entered separately: any general hospital admission ($n = 393$) and any general practice visit ($n = 2660$) in the last 12 months. Results showed that general practice visits had no significant impact on predicting the sensitivity of any of the three instruments. A general hospital admission, however, increased the chance of being detected by the LAST for any criterion group [odds ratio (OR) = 2.03; CI = 1.03–3.99; $P < 0.05$], but not for the single groups. Using the number of doctor visits and the number of hospital admissions as independent variable revealed the following results. The number of doctor visits was related to higher odds ratios in detecting alcohol misuse by the AUDIT-C (OR = 1.95; CI = 1.06–3.58; $P < 0.05$); this relationship was close to significance for the AUDIT ($P = 0.059$). The number of general hospital admissions had an impact on the detection by the LAST (OR = 1.73; CI = 1.05–2.86; $P < 0.05$).

When calculating Cronbach's alpha for the subsample with primary-care utilization in the previous year, internal consistency increased for all questionnaires in those individuals with a general hospital admission in the previous 12 months (AUDIT: from 0.75 to 0.85; AUDIT-C: from 0.56 to 0.66; LAST: from 0.72 to 0.77).

Impact of age and gender on performance

Age showed no meaningful correlations (Spearman rho) with total scores of AUDIT (–0.052), AUDIT-C (–0.037), and LAST (0.015). The mean scores of the questionnaires were compared between male and female participants (t -test). The AUDIT showed higher scores for male (mean \pm SD = 4.49 \pm 3.63) than for female (2.71 \pm 2.02) participants ($t = 18.13$; $df = 2838.85$; $P < 0.001$). The AUDIT-C also had higher scores for male (3.80 \pm 1.62) than for female (2.51 \pm 1.42) participants ($t = 22.70$; $df = 3321.82$; $P < 0.001$). The LAST showed higher scores for male (0.47 \pm 0.103) than for female (0.20 \pm 0.71) individuals ($t = 8.96$; $df = 3215.16$; $P < 0.001$).

As shown in Table 5, areas under the ROC curve differed between men and women. AUDIT and AUDIT-C revealed significantly larger areas for at-risk drinking and any criterion in women. Logistic regression analyses, using gender as dependent variable, revealed differences in sensitivity. Sensitivity was

Table 4. Differences in sensitivity and specificity between tests (McNemar) for adjusted cut-off points

	At-risk drinking ($n = 191$)	Current alcohol misuse ($n = 41$)	Current alcohol dependence ($n = 49$)	Any criterion ^a ($n = 281$)
Sensitivity	AUDIT > AUDIT-C* AUDIT > LAST** AUDIT-C > LAST**	AUDIT > AUDIT-C ^{ns} AUDIT > LAST ^{ns} AUDIT-C > LAST ^{ns}	AUDIT > AUDIT-C ^{ns} AUDIT > LAST ^{ns} AUDIT-C > LAST ^{ns}	AUDIT > AUDIT-C* AUDIT > LAST** AUDIT-C > LAST**
Specificity	AUDIT < AUDIT-C** AUDIT < LAST** AUDIT-C = LAST ^{ns}	AUDIT < AUDIT-C** AUDIT < LAST** AUDIT-C < LAST*	AUDIT > AUDIT-C** AUDIT > LAST** AUDIT-C < LAST*	AUDIT < AUDIT-C** AUDIT < LAST** AUDIT-C = LAST ^{ns}

n (total sample) = 3551.

Cut-off of 5 for AUDIT and AUDIT-C and of 1 for LAST for all subgroups; for current alcohol dependence, the AUDIT was used with 6 as cut-off point.

^aCurrent at-risk drinking, alcohol misuse or dependence.

* $P < 0.05$; ** $P < 0.001$; ns, not significant.

For abbreviations, see Table 1.

Table 5. Gender differences in the performance of AUDIT, AUDIT-C and LAST for different diagnostic groups

Diagnostic group (<i>n</i> : male/female)	Area under ROC curve								
	AUDIT			AUDIT-C			LAST		
	Male	Female	<i>P</i>	Male	Female	<i>P</i>	Male	Female	<i>P</i>
At-risk drinking (122/69)	0.82	0.91	<0.001	0.85	0.91	<0.01	0.62	0.62	1.0
Current alcohol misuse (35/6)	0.77	0.79	0.75	0.73	0.80	0.21	0.67	0.52	0.26
Current alcohol dependence (39/10)	0.94	0.98	0.09	0.90	0.95	0.10	0.86	0.93	0.32
Any criterion ^a (196/85)	0.86	0.91	<0.01	0.86	0.91	<0.01	0.69	0.65	0.38

n (total sample) = 3551.

^aCurrent at-risk drinking, alcohol misuse or dependence.

For abbreviations, see Table 1.

determined by using the adjusted lower cut-offs as described above. The AUDIT was less sensitive in women in detecting at-risk drinking (OR = 0.27; CI = 0.13–0.54; $P < 0.001$), alcohol misuse (OR = 0.09; CI = 0.01–0.88; $P < 0.05$), and any criterion (OR = 0.29; CI = 0.16–0.52; $P < 0.0001$). Results were similar for the AUDIT-C with respect to two diagnostic groups: at-risk drinking (OR = 0.26; CI = 0.13–0.50; $P < 0.001$) and any criterion (OR = 0.30; CI = 0.17–0.53; $P < 0.0001$); results were only close to significance for alcohol misuse ($P = 0.06$). The LAST showed less sensitivity in detecting any criterion group in women (OR = 0.54; CI = 0.32–0.91; $P < 0.05$).

DISCUSSION

To our knowledge, this is the first study providing data on the performance of AUDIT, AUDIT-C and LAST in a general population sample on grounds of diagnostic criteria as gold standard. Our findings suggest that AUDIT and LAST have insufficient sensitivity and AUDIT-C insufficient specificity, when using recommended cut-off scores to detect target groups, i.e. those drinking at at-risk levels, meeting M-CIDI alcohol dependence or misuse criterion, or all three. Only for current alcohol dependence, AUDIT and LAST showed satisfactory sensitivity with standard cut-off points. According to data presented here, the AUDIT performs best with 5 points as a threshold, and the LAST with 1 point. In detecting current alcohol dependence, the AUDIT can be recommended with 6 points as cut-off to achieve higher specificity. For the AUDIT-C, a higher cut-off point of 5 for all target groups is recommended, to improve specificity. Besides these recommendations, it has to be mentioned that choosing a cut-off point depends on the particular purpose and the tolerable rates of false-negatives or false-positives. The findings of this study may serve as a basis to choose the appropriate thresholds for specific purposes.

Comparing the validity measures of the questionnaires with the above-recommended cut-off scores revealed that AUDIT outperforms AUDIT-C and LAST with respect to sensitivity in detecting at-risk drinking. No significant differences in sensitivity between tests could be found for current alcohol dependence or misuse. It must be remembered that the latter groups have smaller sample sizes ($n = 49$ and 41 , respectively), compared to the at-risk drinking group ($n = 191$). Therefore,

differences between tests for alcohol dependence or misuse might reach significance when using larger sample sizes. On the other hand, only small differences in specificity reached significance, because of the respective large groups of individuals with no diagnosis.

The fact that the AUDIT-C, which is a pure alcohol consumption measure, is as good as the full AUDIT, which comprised additional items on negative consequences and signs of dependence, is surprising. However when comparing areas under the ROC curves, there was a tendency ($P = 0.08$) of the AUDIT to perform better in detecting alcohol dependence. Further research is necessary to confirm the performance of the AUDIT-C in detecting alcohol dependence or misuse. It is important to mention that the AUDIT-C was presented as part of the AUDIT in our study; the instrument might have performed differently if the questionnaire had been given without items 4–10 of the AUDIT.

Findings with respect to the performance of the screening questionnaires for individuals who did, and those who did not, report previous health-care utilization are not so clear cut. This corresponds with findings on the detection of alcohol dependence by AUDIT and TWEAK from an US general population sample (Cherpitel, 1998, 1999). In our sample, a previous general hospital admission in the last year increased the chance of being detected by the LAST as having any criterion (at-risk drinking, alcohol misuse or alcohol dependence). Data suggest that the LAST reveals a higher sensitivity in general hospital patients. This is in line with the fact that this instrument was developed in a general hospital setting and incorporates items with the more clinical aspects (liver problems, hospital admission). The number of doctor visits corresponds only with a higher sensitivity in detecting alcohol misuse of the AUDIT-C and (as a tendency) for the AUDIT. The lack of uniform findings with respect to the impact of previous health-care utilization suggests that the relationship is more multi-faceted. One confounding variable might be found in the current influence of the setting on the disclosure of alcohol problems when visiting a doctor or being admitted to a hospital. Findings on an elevated readiness to change drinking behaviour of alcohol-dependent individuals in a general hospital, compared to alcohol-dependent subjects in the general population (Rumpf *et al.*, 1999), underpin this assumption. Hence, retrospective questions on health-care utilization cannot assess all relevant aspects, when simulating the setting specific performance of screening questionnaires in a general population

sample. Assessment at a time other than during admission to a health-care facility may well result in different findings.

In our sample, AUDIT and LAST revealed moderate Cronbach's alpha (AUDIT: 0.75; LAST: 0.72). Although it has to be considered that Cronbach's alpha is related to the number of items, internal consistency for the AUDIT-C was quite poor (alpha = 0.56). The internal consistency for the AUDIT corresponds with the lower part of the range in five studies reviewed by Allen *et al.* (1997). Cronbach's alpha for the AUDIT ranged from 0.75 to 0.94 in different samples including primary care (Barry and Fleming, 1993; Schmidt *et al.*, 1995), college students (Fleming *et al.*, 1991), and individuals arrested for driving while intoxicated (Hays *et al.*, 1995). No data are available on the internal consistency of the AUDIT-C. Cronbach's alpha for the LAST was lower, compared to data from general hospital (0.77 to 0.81) and slightly higher compared to general practice data (0.69) (Rumpf *et al.*, 1997). Using only those individuals with previous general hospital admission in the last 12 months for analysis, increased alpha for all instruments. Our findings suggest that these screening instruments would all show higher reliability in clinical subsamples.

No age-related differences in the performance of the tests were found, but there were significant gender effects. In women, lower mean scores were observed for all three questionnaires, the areas under the ROC curves were larger for at-risk drinking and any criterion using AUDIT and AUDIT-C, and sensitivity was lower for some criterion groups in all instruments. Data suggest that lower cut-offs should be used in female subjects, which is in line with a review on alcohol screening questionnaires in women (Bradley *et al.*, 1998). Moreover, it might be worthwhile to develop gender-specific questionnaires.

Some limitations of the present study have to be considered. The prevalence rates of alcohol-related disorders and at-risk drinking in our sample were quite low, due to drinking practices in the study area involved. The catchment area is in a state near the bottom of the range for Germany which has substantial regional variations in alcohol consumption (Meyer *et al.*, 1998). Although this may have lowered the internal consistency of the questionnaires, we believe this did not substantially affect their validity. However, low prevalence rates have an impact on choosing a cut-off score. People living in areas with high rates of at-risk drinking might have lower scores on some items reflecting social norms of drinking. As a consequence, it might be necessary to change cut-off points. In addition, choosing a cut-off depends on prevalence rates. Given a fixed cut-off point, the probability of a positive screening result being true becomes lower, as the prevalence decreases (positive predictive value).

AUDIT followed by LAST were presented at the end of a comprehensive interview including the substance misuse module of the M-CIDI. Therefore, serial effects have to be considered. One study showed that the sensitivity of the CAGE was lower if alcohol consumption questions were asked first (Steinweg and Worth, 1993). The quantity-frequency assessment of the M-CIDI may have had a similar effect on AUDIT and LAST. However, comprehensive material not related to alcohol was presented between the alcohol use disorders section of the M-CIDI and both questionnaires. Moreover, the vast majority of questions of the entire interview were on mental health and not related to alcohol use. These facts make it rather unlikely that serial effects led to a significant bias in this study.

Our data underline previous findings that screening questionnaires show different validity measures in the general population. Moreover, results presented here suggest that screening instruments are less reliable in the general population, compared to clinical settings. Therefore, data of screening measures in the general population have to be interpreted carefully. To improve the accuracy of screening, the use of two or more complementary instruments should be considered (Rumpf *et al.*, 1998b). Another way of improving screening might be to use simple criteria, such as gender or health-care utilization, to decide which instrument or which cut-off is most adequate. Finally, it is desirable to have a number of different or modified screeners available that perform best in specific settings such as general practice, general hospital, emergency room, work place or the general population.

Acknowledgements — The present data are part of the German research network 'Analytical Epidemiology of Substance Misuse' (ANEPSA). Factors related to the use and misuse of psychoactive substances are analysed by different research groups in the context of several longitudinal studies. For details see ANEPSA Research Group (1998). Contact persons are: G. Bühringer/H. Küfner (IFT Institute for Therapy Research, Munich), H. U. Wittchen/R. Lieb (Max-Planck-Institute, Munich), and U. John (University of Greifswald)/H. Dilling (Medical University of Lübeck). The Research network is funded in the context of the program 'Biological and psycho-social factors of drug misuse and dependence' by the Federal Ministry of Education and Research. Data described in this paper are derived from the project: 'Drug use in the adult general population and remission from drug misuse without formal help', part 1: 'Drug use in the adult general population in a northern German city and surrounding communities'; principal investigators: U. John (University of Greifswald), H. Dilling (Medical University of Lübeck).

REFERENCES

- Allen, J. P., Litten, R. Z., Fertig, J. B. and Babor, T. (1997) A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcoholism: Clinical and Experimental Research* **21**, 613–619.
- American Psychiatric Association (1995) *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn, international version. American Psychiatric Association, Washington, DC.
- ANEPSA Research Group (1998) German research network 'Analytical Epidemiology of Substance Abuse' (ANEPSA). *European Addiction Research* **4**, 203–204.
- Babor, T. F., de la Fuente, J. R., Saunders, J. and Grant, M. (1989a) The alcohol use disorders identification test: guidelines for use in primary health care. In *AUDIT*. World Health Organization, Division of Mental Health, Geneva.
- Babor, T. F., Kranzler, H. R. and Lauerman, R. J. (1989b) Early detection of harmful alcohol consumption: comparison of clinical, laboratory, and self-report screening procedures. *Addictive Behaviors* **14**, 139–157.
- Barry, K. L. and Fleming, M. F. (1993) The Alcohol Use Disorders Identification Test (AUDIT) and the SMAST-13 predictive validity in a rural primary care sample. *Alcohol and Alcoholism* **28**, 33–42.
- Bisson, J., Nadeau, L. and Demers, A. (1999) The validity of the CAGE scale to screen for heavy drinking and drinking problems in a general population survey. *Addiction* **94**, 715–722.
- Bohn, M. J., Babor, T. F. and Kranzler, H. R. (1995) The Alcohol Use Disorders Identification Test (AUDIT): validation of a screening instrument for use in medical settings. *Journal of Studies on Alcohol* **56**, 423–432.
- Bradley, K. A., Boyd-Wickizer, J., Powell, S. H. and Burman, M. L. (1998) Alcohol screening questionnaires in women: a critical review. *Journal of the American Medical Association* **280**, 166–171.
- British Medical Association (1995) *Guidelines on Sensible Drinking*. British Medical Association, London.
- Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D. and Bradley, K. A. (1998) The AUDIT Alcohol Consumption Questions

- (AUDIT-C). An effective brief screening test for problem drinking. *Archives of Internal Medicine* **158**, 1789–1795.
- Chan, A. W. K., Pristach, E. A., Welte, J. W. and Russel, M. (1993) Use of the TWEAK test in screening for alcoholism/heavy drinking in three populations. *Alcoholism: Clinical and Experimental Research* **17**, 1188–1192.
- Chan, A. W., Pristach, E. A. and Welte, J. W. (1994a) Detection by the CAGE of alcoholism or heavy drinking in primary care outpatients and the general population. *Journal of Substance Abuse* **6**, 123–135.
- Chan, A. W., Pristach, E. A. and Welte, J. W. (1994b) Detection of alcoholism in three populations by the Brief-Mast. *Alcoholism: Clinical and Experimental Research* **18**, 695–701.
- Cherpitel, C. J. (1998) Performance of screening instruments for identifying alcohol dependence in the general population, compared with clinical populations. *Alcoholism: Clinical and Experimental Research* **22**, 1399–1404.
- Cherpitel, C. J. (1999) Screening for alcohol problems in the U.S. general population: a comparison of the CAGE and TWEAK by gender, ethnicity, and services utilization. *Journal of Studies on Alcohol* **60**, 705–711.
- Claussen, B. and Aasland, O. G. (1993) The Alcohol Use Disorders Identification Test (AUDIT) in a routine health examination of long-term unemployed. *Addiction* **88**, 363–368.
- Ewing, J. A. (1984) Detecting alcoholism: The CAGE questionnaire. *Journal of the American Medical Association* **252**, 1905–1907.
- Fleming, J. (1996) The epidemiology of alcohol use in Australian women: findings from a national survey of women's drinking. *Addiction* **91**, 1325–1334.
- Fleming, M. F., Barry, K. L. and MacDonald, R. (1991) The Alcohol Use Disorders Identification Test (AUDIT) in a college sample. *International Journal of the Addictions* **26**, 1173–1185.
- Hapke, U., Rumpf, H.-J., Meyer, C., Dilling, H. and John, U. (1998) Substance use, abuse and dependence among the adult population in a rural and urban region of Northern Germany. *European Addiction Research* **4**, 208–209.
- Hays, R. D., Merz, J. F. and Nicholas, R. (1995) Response burden, reliability, and validity of the CAGE, Short MAST, and AUDIT alcohol screening measures. *Behavioral Research Methods, Instruments & Computers* **27**, 277–280.
- Holmila, M. (1995) Intoxication and hazardous use of alcohol: results from the 1992 Finnish Drinking Habits Study. *Addiction* **90**, 785–792.
- Isaacson, J. H., Butler, R., Zacharek, M. and Tzelepis, A. (1994) Screening with the Alcohol Use Disorders Identification Test (AUDIT) in an inner-city population. *Journal of General Internal Medicine* **9**, 550–553.
- Lachner, G., Wittchen, H.-U., Perkonig, A., Holly, A., Schuster, P., Wunderlich, U., Türk, D., Garczynski, E. and Pfister, H. (1998) Structure, content and reliability of the Munich-Composite International Diagnostic Interview (M-CIDI) substance use sections. *European Addiction Research* **4**, 28–41.
- Mayfield, D., McLeod, G. and Hall, P. (1974) The CAGE Questionnaire: validation of a new alcoholism screening instrument. *American Journal of Psychiatry* **131**, 1121–1123.
- McClish, D. K. (1991) Combining and comparing area estimates across studies or strata. *Medical Decision Making* **12**, 274–279.
- Medina-Mora, E., Carreno, S. and De la Fuente, J. R. (1998) Experience with the Alcohol Use Disorders Identification Test (AUDIT) in Mexico. In *Recent Developments in Alcoholism*, Vol. 14: *The Consequences of Alcoholism*, Galanter, M. ed., pp. 383–396. Plenum Press, New York.
- Meyer, C., Rumpf, H.-J., Hapke, U. and John, U. (1998) Regionale Unterschiede in der Prävalenz riskanten Alkoholkonsums: Sekundäranalyse des Gesundheitsurvey Ost-West [Regional differences in the prevalence of hazardous alcohol consumption: a reanalysis of the 'Health Survey East and West Germany']. *Gesundheitswesen* **60**, 486–492.
- Pokorny, A. D., Miller, B. A. and Kaplan, H. B. (1972) The Brief MAST: a shortened version of the Michigan Alcoholism Screening Test. *The American Journal of Psychiatry* **129**, 342–345.
- Robins, L. N., Wing, J. and Wittchen, H. U. (1988) The Composite International Diagnostic Interview: an epidemiological instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry* **45**, 1069–1077.
- Rumpf, H.-J., Hapke, U., Hill, A. and John, U. (1997) Development of a screening questionnaire for the general hospital and general practices. *Alcoholism: Clinical and Experimental Research* **21**, 894–898.
- Rumpf, H.-J., Hapke, U., Dawedeit, A., Meyer, C. and John, U. (1998a) Triggering and maintenance factors of remitting from alcohol dependence without formal help. *European Addiction Research* **4**, 209–210.
- Rumpf, H.-J., Hapke, U., Erfurth, A. and John, U. (1998b) Screening questionnaires in the detection of hazardous alcohol consumption in the general hospital — direct or disguised assessment? *Journal of Studies on Alcohol* **59**, 698–703.
- Rumpf, H.-J., Hapke, U., Meyer, C. and John, U. (1999) Motivation to change drinking behavior: comparison of alcohol-dependent individuals in a general hospital and a general population sample. *General Hospital Psychiatry* **21**, 348–353.
- Russell, M., Martier, S. S., Sokol, R. J., Mudar, P., Bottoms, S., Jacobson, S. and Jacobson, J. (1994) Screening for pregnancy risk-drinking. *Alcoholism: Clinical and Experimental Research* **18**, 1156–1161.
- Saunders, J. B., Aasland, O. G., Babor, T. F., DeLaFuente, J. R. and Grant, M. (1993) Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption — II. *Addiction* **88**, 617–629.
- Schmidt, A., Barry, K. L. and Fleming, M. F. (1995) Detection of problem drinkers: The Alcohol Use Disorders Identification Test (AUDIT). *Southern Medical Journal* **88**, 52–59.
- Selzer, M. L. (1971) The Michigan Alcoholism Screening Test: the quest for a new diagnostic instrument. *American Journal of Psychiatry* **127**, 1653–1658.
- Selzer, M. L., Vinokur, A. and Rooijen, M. A. (1975) A self-administered Short Michigan Alcoholism Screening Test (SMAST). *Journal of Studies on Alcohol* **36**, 117–126.
- Steinweg, D. L. and Worth, H. (1993) Alcoholism: the keys to the CAGE. *American Journal of Medicine* **94**, 520–523.
- Wittchen, H.-U., Beloch, E., Garczynski, E., Holly, A., Lachner, G., Perkonig, A., Vodermaier, A., Vossen, A., Wunderlich, U. and Ziegglänsberger, S. (1995) *Münchener Composite International Diagnostic Interview (M-CIDI)*, Version 2.2. Max-Planck-Institut für Psychiatrie, München.
- Wittchen, H.-U., Lachner, G., Wunderlich, U. and Pfister, H. (1998) Test-retest reliability of the computerized DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). *Social Psychiatry and Psychiatric Epidemiology* **33**, 568–578.