The alcohol use disorders identification test for detecting at-risk drinking: a systematic review and meta-analysis

Berner MM, Kriston L, Bentele M, Huter M

CRD summary
This review assessed the diagnostic accuracy of the Alcohol Use Disorders Identification Test (AUDIT) to detect at-risk drinking. It concluded that the AUDIT should only be used for primary care populations, hospital in-patients, and elderly patients. Given the limitations of the analysis and of the methods of the included studies, plus the variability between studies, this conclusion may be overstated.

Authors' objectives
To determine the diagnostic accuracy of the Alcohol Use Disorders Identification Test (AUDIT) to detect at-risk drinking.

Searching
MEDLINE, PsycINFO, CINAHL, Science Citation Index Expanded, BIOSIS previews, Medion, and the Cochrane Library were searched for articles from January 1989 to November 2004. Search terms were reported and there were no language restrictions. The reference lists of relevant articles were also checked.

Study selection
Studies were considered eligible for inclusion if they reported the full 10-point AUDIT scores for each participant; compared AUDIT scores with a valid reference standard for at-risk alcohol consumption and drinking patterns for all participants; and did not use all or part of the AUDIT scores as part of the reference standard; and if the reference test was performed within one month of the AUDIT and on at least half of the participants.

Where reported, the mean age of the participants ranged from 20.3 years to 77.1 years and most of the studies were conducted in the USA, the UK, or Spain and in a primary care setting. The threshold for the AUDIT ranged across studies from three to ten. The reference standard threshold varied widely, including between two and three drinks a day, between seven and 28 drinks a week, between 14 and 21 units a week, and between 20 and 57 drinks in the last month or four weeks. In some studies the reference levels for males and females were the same and in others they were not. In a substantial minority of studies, participants with more serious alcohol abuse disorders were included in the sample.

One reviewer screened titles and abstracts to assess whether studies might be eligible for inclusion; the full texts of the identified papers were then assessed independently by two other reviewers. Disagreements were resolved by consensus.

Assessment of study quality
The quality of studies was assessed using criteria from the Quality Assessment of Diagnostic Accuracy Studies checklist. This included how representative the sample was of the patient population; the clarity of patient selection criteria; the avoidance of partial verification bias; sufficient description of the execution of reference measures; blinding to allocation, of the interpreters of the tests (assessed as a single criterion); and explanation for withdrawals. Studies had to avoid differential verification, incorporation, and progression biases to be included.

It appears that at least two reviewers conducted the quality assessment and disagreements were resolved by consensus.

Data extraction
Two reviewers independently extracted the data required to calculate the sensitivity, specificity, positive and negative likelihood ratios, and diagnostic odds ratios, along with 95% confidence intervals. Where possible, these were calculated for two AUDIT score thresholds: eight or over, as recommended by the developers of the AUDIT (general); and thresholds that were considered appropriate by the authors for the study population (specific).
Methods of synthesis

Pooled likelihood ratios and diagnostic odds ratios were calculated using a random-effects model for both general and specific thresholds, irrespective of heterogeneity. Heterogeneity was assessed, using the Cochran Q and $I^2$ tests. Summary estimates of sensitivity and specificity were calculated, using weighted averages.

Studies were subgrouped into categories that included general primary care; general medical in-patients; emergency department or trauma centre; adolescents or college students; elderly patients; patients with mental disorders; and highly specific populations. Where a subgroup included at least five studies, a summary receiver operating characteristic curve was produced, using the Moses-Littenberg model, and the area under the curve was calculated. The potential for a threshold effect was investigated using $\beta$ values of the summary receiver operating characteristic curve and Spearman’s rank-correlation coefficients.

Funnel plots were visually inspected to assess publication bias.

Results of the review

Twenty-three studies were included in the review, and 19, with over 26,000 participants, were meta-analysed. Of the included studies, 59% recruited a representative patient sample, 86% avoided partial verification bias and adequately described the execution of the reference standard, 27% blinded the interpreters of tests, and 77% explained withdrawals and clearly described the selection criteria. Overall, 55% of studies were considered to be of high quality.

Total pooled estimates of sensitivity and specificity were not calculated, due to statistically significant heterogeneity ($I^2 \geq 83\%$). None of the planned subgroups substantially reduced the heterogeneity, except for specificity for elderly patients at the general threshold (0.96, 95% CI 0.95 to 0.98; three studies), which was zero.

With the general AUDIT threshold (17 studies), the pooled positive likelihood ratio was 7.18 (95% CI 5.44 to 9.48), the negative likelihood ratio was 0.34 (95% CI 0.25 to 0.46), and the diagnostic odds ratio was 26.29 (95% CI 17.23 to 40.14). From the summary receiver operating characteristic analysis of the eight primary care studies, the area under the curve was 0.92 (95% CI 0.87 to 0.96). A funnel plot, for primary care studies, indicated no evidence of publication bias.

With specific AUDIT thresholds (16 studies), the pooled positive likelihood ratio was 5.38 (95% CI 4.35 to 6.65), the negative likelihood ratio was 0.21 (95% CI 0.16 to 0.28), and the diagnostic odds ratio was 28.65 (95% CI 18.90 to 43.42).

Authors’ conclusions

The results reported were highly heterogeneous, and the AUDIT should only be used with primary care populations, hospital in-patients, or elderly patients.

CRD commentary

This review addressed a clear question with well-defined inclusion criteria. Several appropriate sources were searched, without language restrictions. No attempts were made to identify unpublished studies and studies could have been missed. Titles and abstracts were assessed by one reviewer, which means that reviewer error and bias cannot be ruled out. All other stages of the review were conducted in duplicate. Study quality was assessed, using appropriate criteria, and the decision not to produce pooled estimates of sensitivity and specificity was appropriate, given the heterogeneity across studies and the evidence of a threshold effect on the one summary receiver operating characteristic analysis that was conducted. The results of the investigation into this potential threshold effect were not reported. Substantial heterogeneity was present for most of the analyses of the other diagnostic outcomes and so the reliability and generalisability of these pooled estimates is uncertain. The summary receiver operating characteristic model did not take into account the between-study variability and it was difficult to match the points on the curve to the values in the results table, to determine which studies contributed to this analysis.

The review process was generally well-conducted, but there were limitations. These and the limitations in the methods of some of the studies, plus the clinical and statistical heterogeneity between them, mean that the conclusion and
implications for practice might be overstated.

**Implications of the review for practice and research**

**Practice:** The authors stated that, until more research is available, the AUDIT should only be used with a recommended threshold, obtained in a primary study of a population and setting similar to the target population and setting.

**Research:** The authors stated that there should be further research into the AUDIT, using shorter versions of the test, and into the acceptability and feasibility of early intervention in different health care settings.

**Funding**
German Federal Ministry of Education and Research (BMBF), grant number 01EB0111.

**Bibliographic details**

**PubMedID**
17446987

**Original Paper URL**
http://www.jsad.com/jsad/article/The_Alcohol_Use_Disorders_Identification_Test_for_Detecting_AtRisk_Drinkin/2141.html

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Alcohol Drinking /adverse effects; Alcohol-Related Disorders /diagnosis /psychology; Humans; Mass Screening; Personality Inventory /statistics & numerical data; Psychometrics /statistics & numerical data; Reproducibility of Results; Risk Assessment; Surveys and Questionnaires

**AccessionNumber**
12007001755

**Date bibliographic record published**
07/01/2008

**Date abstract record published**
11/08/2010

**Record Status**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.