Accuracy of one or two simple questions to identify alcohol-use disorder in primary care: a meta-analysis

INTRODUCTION
Alcohol problems are common in primary care with between 7% and 30% of attendees having at-risk drinking or an alcohol-use disorder.1–3 In a meta-analysis of five in primary care studies the pooled prevalence of lifetime problem drinking was 19.8%.4 Many authors have therefore suggested that GPs are well positioned to identify and manage alcohol problems.5–9 Several effective treatment packages called ‘brief alcohol interventions’ have been developed and tested.10,11 Clearly these interventions require adequate initial identification of alcohol problems, yet only about one-third of such individuals are detected by their GP.12–14 Studies conducted in the US, UK, Australia, and Finland indicate that clinicians frequently do not screen for problem drinking, and fail to address the problem even when the diagnosis is known.15–18 A recent meta-analysis found that GPs had a clinical detection sensitivity of 42% but alcohol problems were recorded correctly in the notes only 27.3% of the time.19

In response to these concerns The Institute of Medicine, the National Institute on Alcohol Abuse and Alcoholism (NIAAA), the American Medical Association, and the American Society of Addiction Medicine have all recommended that clinicians routinely ask patients about alcohol use.20–23 In order to screen for problem drinking, the US NIAAA and the US Preventive Services Task Force (USPSTF) recommend population screening; that is, physicians should ask all attendees whether they drink, and assess the specific quantity, frequency, and pattern of consumption, but they did not recommend a specific tool.24 In 2004, the USPSTF recommended screening is accompanied by behavioural counselling interventions to reduce alcohol misuse by adults in primary care settings.25 The NIAAA also recommends targeted screening (case-finding) in that all patients who drink alcohol should be screened with the four CAGE questions.26 In the UK, the Primary Care Service Framework prefers the 10 AUDIT questions; in fact, it recommends an algorithm approach whereby either the Single Alcohol Screening Question (SASQ) (1 item), Fast Alcohol Screening Test (FAST) (4 items), Alcohol-use disorders Identification Test (AUDIT)-C (3 items) or AUDIT-PC (5 items) is applied as an initial first step, followed by the remaining AUDIT questions given to those who initially score positive on the screening test.27 Yet choice
of the initial steps is not yet clear, and no evidence has yet been presented for algorithm approaches.

In reality, many GPs have difficulty applying routine alcohol screening, although most state that they often inquire about drinking behaviours. Indeed observational studies have shown that screening for alcohol problems is far from routine in primary care. In the UK, alcohol screening sometimes occurs at patient registration but is usually opportunistic. Investigations into alcohol screening practices have consistently found that most physicians ask patients about consumption, but few go beyond an initial inquiry. Most alcohol discussions last less than 1 minute. D’Amico et al examined primary care physician practices in over 7000 visits; practitioners asked only 29% of all attendees about their drinking over the course of 1 year, although they asked 44% of problem drinkers about their drinking. Of those individuals with problem drinking, 49% received relevant advice.

Lessons from screening studies for depression suggest that ultra-short methods consisting of one or two questions might be valuable in some circumstances. However such methods often have low positive predictive value. The aim of this study was to find out whether very simple one and two question screening might prove an accurate and acceptable screening method in primary care using quantitative meta-analysis. It was hypothesised that one question might have good rule-in but poor rule-out performance when used alone.

**METHODS**

**Inclusion/exclusion criteria**
The principal inclusion criteria were studies that examined the diagnostic accuracy of simple verbal questions to identify those with defined alcohol problems in primary care. No language restrictions were applied.

This study was principally concerned with the accuracy of detecting those individuals with defined alcohol problems, as opposed to an assessment of alcohol intake, therefore to minimise heterogeneity, the criterion standard required (established by interview) was the DSM or ICD ‘alcohol-use disorder’, rather than alcohol dependence. Alcohol-use disorder can be further subdivided into alcohol abuse and alcohol dependence. Studies that did not present sufficient primary data or failed to employ an adequate criterion standard were excluded (such as studies based on self-reported alcohol use). The initial citations were screened to exclude studies that did not assess 1- or 2-item questions to identify alcohol problems. After this initial sifting process, papers were independently evaluated for eligibility. Disagreements were resolved by discussion.

**Search, sources, and appraisal**
A systematic literature search, critical appraisal and meta-analysis were conducted. MEDLINE/PubMed and Embase abstract databases were searched from inception to January 2014 (Appendix 1). In four full text collections (Science Direct, Ingenta Select, Ovid Full text, Blackwell-Wiley Interscience) the same search terms were used (Appendix 1) but as a full text search and citation search. The abstract database, Web of Knowledge (4.0, ISI) was searched, using the same terms as in Appendix 1 as a text word search, and using key papers in a reverse citation search. Appraisal of each article was conducted by all authors independently using QUADAS. This is a standardised quality appraisal form and is recommended by a number of organisations such as the Cochrane Collaboration and the National Institute of Health and Clinical Excellence. No abstracts or non-English articles were found.

**Data collection process**
A standardised data extraction form was used, which was developed and piloted on several previous systematic reviews of diagnostic accuracy conducted by the authors. Data extraction was conducted independently, checked, and differences were resolved by discussion. Variables extracted were: country of study, setting, patient characteristics (such as age and sex), reference standard (including cut-off if relevant), question used to identify alcohol problems, sample size, blindness of the interviewers to the index test, and Youden scores (sensitivity + specificity − 1). For the purposes of the meta-analyses sensitivity,
specificity, and prevalence of alcohol-use disorder (as measured by the reference standard) were extracted. In addition, if not provided in the papers, 2x2 tables (true positives, false positives, true negatives, and false negatives) were calculated from this data for inclusion in the meta-analysis. Secondary outcomes were an area under the curve analysis (see below) for screening and case-finding performance.

**Analysis**

**Meta-analysis.** Random effects meta-analysis were used to synthesise and provide pooled estimates of sensitivity and specificity for the use of a single question or two questions to identify people with alcohol problems. In addition, a bivariate meta-analysis was conducted using the metandi commands in Stata (version 10). This method fits a 2-level model, with independent binomial distributions for the true positives and true negatives conditional on the sensitivity and specificity in each study, and a bivariate normal model for the logit transforms of sensitivity and specificity between studies. A summary Receiver Operator Characteristic curve (sROC), where each data point represents a separate study, was then constructed using the bivariate model to produce a 95% confidence ellipse within ROC space. Additionally, a Bayesian curve analysis was undertaken, which plots all post-test probabilities from all pre-test probabilities regardless of prevalence. Heterogeneity was assessed using the $I^2$ statistic. In addition, publication bias was assessed both visually using funnel plots and formally using the Begg-Mazumdar test.

**Accuracy and clinical utility.** Further analyses were conducted to assess the accuracy and clinical utility of one or two questions to identify people with alcohol problems, both generally (using any one-question or two-question approach) and in comparison with one another. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood ratios and clinical utility index were calculated using the online calculator (www.clinicalutility.co.uk). Clinical utility index allows a qualitative measure of clinical accuracy. Positive clinical utility is a proxy for rule-in accuracy or case-finding. Negative clinical utility is a proxy for rule-out accuracy or screening.

**RESULTS**

**Study description and methods**

There were 15 tests of individual questions identified in six publications and two tests of two-question approaches, involving 5646 unique individuals. All were diagnostic accuracy studies. The sample size of individual studies ranged from 227 to

<table>
<thead>
<tr>
<th>Table 1. QUADAS Methodological appraisal of single-question studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study ID</strong></td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td><strong>Single-question studies</strong></td>
</tr>
<tr>
<td>Bradley et al, 2001</td>
</tr>
<tr>
<td>Brown et al, 2001</td>
</tr>
<tr>
<td>Fleming et al, 1991</td>
</tr>
<tr>
<td>Seale et al, 2006</td>
</tr>
<tr>
<td>Smith et al, 2009</td>
</tr>
<tr>
<td>Volk, et al, 1997</td>
</tr>
<tr>
<td><strong>Two-question studies</strong></td>
</tr>
<tr>
<td>Brown et al, 2001</td>
</tr>
<tr>
<td>Vinson et al, 2007</td>
</tr>
</tbody>
</table>
Table 2. Statistical appraisal of single- and two-question studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Phrase of questions</th>
<th>Criterion standard</th>
<th>Total sample</th>
<th>All cases</th>
<th>Non-cases</th>
<th>Unadjusted sensitivity</th>
<th>Unadjusted specificity</th>
<th>Unadjusted PPV</th>
<th>Unadjusted NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single-question studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seale et al, 2006[1]</td>
<td>SASQ (When was the last time you had more than X drinks in one day?)</td>
<td>Alcohol-use disorder (DSM-IV)</td>
<td>625</td>
<td>127</td>
<td>498</td>
<td>0.866</td>
<td>0.490</td>
<td>0.302</td>
<td>0.935</td>
</tr>
<tr>
<td>Fleming et al, 1991[2]</td>
<td>Unique — Have you ever had a drinking problem?</td>
<td>Alcohol-use disorder (DSM-III)</td>
<td>280</td>
<td>82</td>
<td>198</td>
<td>0.573</td>
<td>0.879</td>
<td>0.662</td>
<td>0.833</td>
</tr>
<tr>
<td>Brown et al, 2001[3]</td>
<td>Unique — Do you have a current drinking problem?</td>
<td>Alcohol-use disorder (DSM-III)</td>
<td>280</td>
<td>82</td>
<td>198</td>
<td>0.317</td>
<td>0.970</td>
<td>0.813</td>
<td>0.774</td>
</tr>
<tr>
<td></td>
<td>Unique — In the last year, how many times have you not remembered things that happened while you were drinking or using drugs?</td>
<td>Substance Use Disorder (CIDI-SAM for DSM-III-R)</td>
<td>1136</td>
<td>261</td>
<td>875</td>
<td>0.483</td>
<td>0.925</td>
<td>0.656</td>
<td>0.857</td>
</tr>
<tr>
<td></td>
<td>Unique — In the last year, have you ever drunk or used drugs more than you meant to?</td>
<td>Substance Use Disorder (CIDI-SAM for DSM-III-R)</td>
<td>1136</td>
<td>261</td>
<td>875</td>
<td>0.701</td>
<td>0.809</td>
<td>0.523</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td>Unique — Have you felt you wanted or needed to cut down on your drinking or drug use in the last year?</td>
<td>Substance Use Disorder (CIDI-SAM for DSM-III-R)</td>
<td>1136</td>
<td>261</td>
<td>875</td>
<td>0.563</td>
<td>0.917</td>
<td>0.668</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td>Unique — In the last year, have you drunk or used non-prescription drugs to deal with your feelings, stress, or frustration?</td>
<td>Substance Use Disorder (CIDI-SAM for DSM-III-R)</td>
<td>1136</td>
<td>261</td>
<td>875</td>
<td>0.544</td>
<td>0.869</td>
<td>0.553</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>Unique — As a result of your drinking or drug use, did anything happen in the last year that you wish didn’t happen?</td>
<td>Substance Use Disorder (CIDI-SAM for DSM-III-R)</td>
<td>1136</td>
<td>261</td>
<td>875</td>
<td>0.410</td>
<td>0.960</td>
<td>0.754</td>
<td>0.845</td>
</tr>
<tr>
<td>Bradley et al, 2001[4]</td>
<td>SASQ-Modified How often do you have six or more drinks on one occasion?</td>
<td>Alcohol-use disorder (DSM-III-R) or hazardous drinking (NAAA)</td>
<td>227</td>
<td>92</td>
<td>135</td>
<td>0.772</td>
<td>0.830</td>
<td>0.755</td>
<td>0.842</td>
</tr>
<tr>
<td></td>
<td>SASQ-Modified Maximum number of drinks per occasion</td>
<td>Alcohol-use disorder (DSM-III-R) or hazardous drinking (NAAA)</td>
<td>227</td>
<td>92</td>
<td>135</td>
<td>0.587</td>
<td>0.770</td>
<td>0.635</td>
<td>0.732</td>
</tr>
<tr>
<td>Volk et al, 1997[5]</td>
<td>CAGE: cut down question</td>
<td>Alcohol-use disorder (DSM-IV)</td>
<td>1333</td>
<td>205</td>
<td>1128</td>
<td>0.629</td>
<td>0.840</td>
<td>0.417</td>
<td>0.926</td>
</tr>
<tr>
<td></td>
<td>CAGE: annoyed question</td>
<td>Alcohol-use disorder (DSM-IV)</td>
<td>1333</td>
<td>205</td>
<td>1128</td>
<td>0.239</td>
<td>0.930</td>
<td>0.383</td>
<td>0.871</td>
</tr>
<tr>
<td></td>
<td>CAGE: guilty question</td>
<td>Alcohol-use disorder (DSM-IV)</td>
<td>1333</td>
<td>205</td>
<td>1128</td>
<td>0.371</td>
<td>0.900</td>
<td>0.402</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>CAGE: eye-opener question</td>
<td>Alcohol-use disorder (DSM-IV)</td>
<td>1333</td>
<td>205</td>
<td>1128</td>
<td>0.210</td>
<td>0.950</td>
<td>0.434</td>
<td>0.869</td>
</tr>
<tr>
<td>Smith et al, 2009[6]</td>
<td>SASQ-Modified How many times in the past year have you had X or more drinks in a day?</td>
<td>Alcohol-use disorder (DSM-IV)</td>
<td>286</td>
<td>33</td>
<td>253</td>
<td>0.879</td>
<td>0.672</td>
<td>0.259</td>
<td>0.977</td>
</tr>
<tr>
<td><strong>Two-question studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown et al, 2001[7]</td>
<td>In the last year, have you ever drunk or used drugs more than you meant to? OR Have you felt you wanted or needed to cut down on your drinking or drug use in the last year?</td>
<td>Substance Use Disorder (CIDI-SAM for DSM-III-R)</td>
<td>1136</td>
<td>261</td>
<td>875</td>
<td>0.793</td>
<td>0.779</td>
<td>0.518</td>
<td>0.927</td>
</tr>
<tr>
<td>Vinson et al, 2007[8]</td>
<td>Q1 Recurrent drinking in situations in which it is physically hazardous, and Q2 drinking in larger amounts or over a longer period than intended</td>
<td>Alcohol-use disorder (DSM-IV)</td>
<td>623</td>
<td>126</td>
<td>496</td>
<td>0.944</td>
<td>0.819</td>
<td>0.569</td>
<td>0.983</td>
</tr>
</tbody>
</table>

NPV = negative predictive value. PPV = positive predictive value.
1333 individuals (33–261 with alcohol-use disorders) (Tables 1 and 2). Most studies used a DSM-based definition of alcohol-use disorder, specifically alcohol abuse and no study examined alcohol dependence. One study defined substance abuse according to DSM-III-R.70 The pooled prevalence of alcohol-use disorder was 21.0% (95% CI = 20.0% to 22.1%).

**Diagnostic accuracy of the single-question test**

Across 15 analyses a single-question approach facilitated the identification of 453 out of 800 unique individuals with problem drinking (cases), an uncorrected pooled sensitivity of 56.6%. There was no evidence of publication bias (Begg-Mazumdar: Kendall’s Tau-b = 0.21, P = 0.29). A single question allowed identification of 2510 or 3087 unique individuals without problem drinking (non-cases), an uncorrected pooled specificity of 81.3%. There was high heterogeneity present (I² = 96.3%). On bivariate meta-analysis sensitivity was adjusted to 54.5% (95% CI = 43.0% to 65.5%) and specificity was 87.3% (95% CI = 81.5% to 91.5%) (Figure 1). At the stated prevalence, the PPV was 53.3% (95% CI = 48.0% to 58.5%) and the NPV was 85.9% (95% CI = 89.5%). Positive likelihood ratio was 4.29 (CI = 3.84 to 4.80) and negative likelihood ratio 0.52 (95% CI = 0.48 to 0.56). The positive clinical utility index score for the single-question test was 0.287 (95% CI = 0.286 to 0.288) rated as ‘very poor’. The negative clinical utility index score was 0.769 (95% CI = 0.769 to 0.769) rated as ‘good’.

The most accurate single method was the AUDIT followed by the CAGE, then two-questions and finally a single question. In terms of algorithm approaches the optimal

**Algorithm approach versus routine screening**

The merits of an algorithm approach similar to that recommended by the Primary Care Service Framework were examined. An algorithm potentially saves time because only those individuals who screen positive after the first step receive the longer instrument. Using primary care data from two previous meta-analyses a combination of one or two brief questions was compared with either the 4-item CAGE (n = 10, pooled sensitivity = 71% specificity = 91%) or the 10-item AUDIT (n = 6, pooled sensitivity = 61.9%, specificity 91.5%). Then the accuracy of an algorithm approach was examined, namely a single question or two questions followed by either the CAGE or AUDIT (Table 3).

The most accurate single method was the AUDIT followed by the CAGE, then two-questions and finally a single question. In terms of algorithm approaches the optimal

![Figure 1. Bayesian pre-test post-test accuracy of single-question approach to alcohol-use disorder in primary care (a = unadjusted; b adjusted for unique patients).](image)
The combination was 2QQ then CAGE (0.909, 95% CI = 0.891 to 0.926) although this was almost identically accurate to the 2QQ then the AUDIT (0.894, 95% CI = 0.875 to 0.913). These approaches would be more efficient than CAGE or AUDIT alone. For example for the assessment of 100 hypothetical primary care attendees the AUDIT alone would need 1000 questions to be asked (10 per subject) but the 2QQ and AUDIT algorithm would require 536.

**Real-world value**

In the same 100 hypothetical attendee scenario as above, at a prevalence of 21%, a GP using only a single question for alcohol-use disorder would correctly identify 11 cases, missing 10. They would correctly identify 69 non-cases and falsely diagnose 10; the overall accuracy would be 80.4%. Using an algorithm approach at the same prevalence, a GP looking for alcohol-use disorder using an algorithm of a single question followed by the CAGE would correctly identify 8 cases, missing 13. They would correctly identify 78 non-cases and falsely diagnose one; the overall accuracy would be 86.1%.

**DISCUSSION**

This study is concerned with the methods of detecting those individuals with defined alcohol problems as opposed to an assessment of alcohol intake. In this meta-analysis of 17 tests from seven publications designed to help with the identification of alcohol-use disorder in primary care, a single question had a sensitivity of only 54.5% [95% CI = 43.0% to 65.5%] and specificity of 87.3% [95% CI = 81.5% to 91.5%] but two questions had a sensitivity of 87.2% [95% CI = 69.9% to 97.7%] and a specificity of 79.8% [95% CI = 75.7% to 83.6%].

**Real-world value**

In the same 100 hypothetical attendee scenario as above, at a prevalence of 21%, a GP using only a single question for alcohol-use disorder would correctly identify 11 cases, missing 10. They would correctly identify 69 non-cases and falsely diagnose 10; the overall accuracy would be 80.4%. Using an algorithm approach at the same prevalence, a GP looking for alcohol-use disorder using an algorithm of a single question followed by the CAGE would correctly identify 8 cases, missing 13. They would correctly identify 78 non-cases and falsely diagnose one; the overall accuracy would be 86.1%.

**DISCUSSION**

This study is concerned with the methods of detecting those individuals with defined alcohol problems as opposed to an assessment of alcohol intake. In this meta-analysis of 17 tests from seven publications designed to help with the identification of alcohol-use disorder in primary care, a single question had a sensitivity of only 54.5% [95% CI = 43.0% to 65.5%] and specificity of 87.3% [95% CI = 81.5% to 91.5%] but two questions had a sensitivity of 87.2% [95% CI = 69.9% to 97.7%] and a specificity of 79.8% [95% CI = 75.7% to 83.6%].

**Real-world value**

In the same 100 hypothetical attendee scenario as above, at a prevalence of 21%, a GP using only a single question for alcohol-use disorder would correctly identify 11 cases, missing 10. They would correctly identify 69 non-cases and falsely diagnose 10; the overall accuracy would be 80.4%. Using an algorithm approach at the same prevalence, a GP looking for alcohol-use disorder using an algorithm of a single question followed by the CAGE would correctly identify 8 cases, missing 13. They would correctly identify 78 non-cases and falsely diagnose one; the overall accuracy would be 86.1%.

**DISCUSSION**

This study is concerned with the methods of detecting those individuals with defined alcohol problems as opposed to an assessment of alcohol intake. In this meta-analysis of 17 tests from seven publications designed to help with the identification of alcohol-use disorder in primary care, a single question had a sensitivity of only 54.5% [95% CI = 43.0% to 65.5%] and specificity of 87.3% [95% CI = 81.5% to 91.5%] but two questions had a sensitivity of 87.2% [95% CI = 69.9% to 97.7%] and a specificity of 79.8% [95% CI = 75.7% to 83.6%].

**Real-world value**

In the same 100 hypothetical attendee scenario as above, at a prevalence of 21%, a GP using only a single question for alcohol-use disorder would correctly identify 11 cases, missing 10. They would correctly identify 69 non-cases and falsely diagnose 10; the overall accuracy would be 80.4%. Using an algorithm approach at the same prevalence, a GP looking for alcohol-use disorder using an algorithm of a single question followed by the CAGE would correctly identify 8 cases, missing 13. They would correctly identify 78 non-cases and falsely diagnose one; the overall accuracy would be 86.1%.

**DISCUSSION**

This study is concerned with the methods of detecting those individuals with defined alcohol problems as opposed to an assessment of alcohol intake. In this meta-analysis of 17 tests from seven publications designed to help with the identification of alcohol-use disorder in primary care, a single question had a sensitivity of only 54.5% [95% CI = 43.0% to 65.5%] and specificity of 87.3% [95% CI = 81.5% to 91.5%] but two questions had a sensitivity of 87.2% [95% CI = 69.9% to 97.7%] and a specificity of 79.8% [95% CI = 75.7% to 83.6%].

**Real-world value**

In the same 100 hypothetical attendee scenario as above, at a prevalence of 21%, a GP using only a single question for alcohol-use disorder would correctly identify 11 cases, missing 10. They would correctly identify 69 non-cases and falsely diagnose 10; the overall accuracy would be 80.4%. Using an algorithm approach at the same prevalence, a GP looking for alcohol-use disorder using an algorithm of a single question followed by the CAGE would correctly identify 8 cases, missing 13. They would correctly identify 78 non-cases and falsely diagnose one; the overall accuracy would be 86.1%.

**DISCUSSION**

This study is concerned with the methods of detecting those individuals with defined alcohol problems as opposed to an assessment of alcohol intake. In this meta-analysis of 17 tests from seven publications designed to help with the identification of alcohol-use disorder in primary care, a single question had a sensitivity of only 54.5% [95% CI = 43.0% to 65.5%] and specificity of 87.3% [95% CI = 81.5% to 91.5%] but two questions had a sensitivity of 87.2% [95% CI = 69.9% to 97.7%] and a specificity of 79.8% [95% CI = 75.7% to 83.6%].

**Real-world value**

In the same 100 hypothetical attendee scenario as above, at a prevalence of 21%, a GP using only a single question for alcohol-use disorder would correctly identify 11 cases, missing 10. They would correctly identify 69 non-cases and falsely diagnose 10; the overall accuracy would be 80.4%. Using an algorithm approach at the same prevalence, a GP looking for alcohol-use disorder using an algorithm of a single question followed by the CAGE would correctly identify 8 cases, missing 13. They would correctly identify 78 non-cases and falsely diagnose one; the overall accuracy would be 86.1%.
Figure 2. Meta-analytic comparison of single question and two question versus CAGE versus AUDIT in primary care (Conditional probability plot using comparison data from CAGE and AUDIT).

**Strengths and limitations**

This study has several limitations. Despite an extensive search only a small number of studies were found that adhered to a robust interview criterion standard. The study was limited by a lack of data regarding ‘at-risk drinking’ and ‘alcohol dependence’. This could be important as some reports suggest that the CAGE questions perform better in identifying alcohol dependence whereas AUDIT is more sensitive for hazardous and harmful drinkers. It was also not possible to separate results by sex, although screening differences may exist in clinical practice for males and females. These issues could be addressed in future research. The study also found considerable statistical heterogeneity which could relate to a wide variety of single-question approaches.

**Comparison with existing literature**

The merits of algorithmic screening have been previously explored. The exact approach recommended by the Primary Care Service Framework, namely the SASQ single question ‘How often do you have six or more drinks on one occasion?’ followed by the AUDIT, was reasonably successful. Only two questions followed by the AUDIT and two questions followed by the CAGE were more accurate but at a cost of an additional 2.3 and 0.2 questions per attendee. In clinical practice a GP using only a single question for alcohol-use disorder would be likely to miss 10 cases and falsely diagnose 10 non-cases for every 100 presentations. Using a single question followed by the CAGE the GP would be likely to miss up to 13 cases but with only one false positive. If this rate of missed cases is considered too high, currently the only viable solution is to use a longer questionnaire for all attendees. However the most common questionnaires for alcohol problems (that is the AUDIT, CAGE, and MAST) are usually considered difficult to use in primary care. Clinicians prepared to use a few more than two questions have a choice of the three question AUDIT-C or the 4-item CAGE, and the 4-item FAST. Of these the FAST and CAGE take only 12–15 seconds in experienced hands but their acceptability in primary care is not yet clear. The FAST is of particular interest because item 1 on the FAST is in fact the SASQ question (as above) and the authors recommend an algorithmic approach whereby low and high scorers to question 1 receive no further screening but those with intermediate scores receive three follow-up questions extracted from the AUDIT. However the FAST has yet to be evaluated against an interview-based criterion standard in primary care, therefore the authors suggest it requires further testing, particularly alongside other approaches. Further, no single laboratory test has been shown to be a substitute for questionnaire-based screening.

**Routine screening**

Experts have recommended routine alcohol screening focusing on new patient registrations, general health checks, and special types of consultation. Mitchell et al previously demonstrated that depression diagnostic
accuracy can be improved with multiple assessments.46 The authors expect the authors’ group to occur in relation to alcohol assessment because patients are often seen on multiple occasions. A recent meta-analysis found that GPs had a clinical detection sensitivity for alcohol problems of 42% but a sensitivity of 93.1% when unassisted.19 These figures could be higher in some countries where the recognition of alcohol use is linked to primary care performance incentives. At face value the single-question performance of 54.5% sensitivity and 87.3% specificity is not appreciably different from unassisted ability. However, two questions with a sensitivity of 87.2% and a specificity of 79.8% does seem to be an improvement. This is similar for ultra-short methods of screening for depression where one question is inadequate but two an improvement over clinician judgement.18 Algorithm approaches are better still and can be efficient at the same time. Here the optimal approach appears to be two questions followed by CAGE, which achieved an overall accuracy of 90.9% and requires only 3.3 questions per attendee.

Acceptability: The clinical value of screening tests depends not only on accuracy but also acceptability. The primary purpose of the single item is as a first-step screen to rule-out those unlikely to have an alcohol problem. The authors suggest it is feasible to require one question to be asked routinely, but it is appropriate to examine acceptability even for a single question, as some questions may not be welcome in unselected primary care attendees. Vinson and colleagues rated patients’ comfort with single-item screening. On an ordinal scale in which 1 is ‘very uncomfortable’ and 5 is ‘very comfortable’, patients rated their comfort as 4.2 with a single-question approach.18 Regarding acceptability for staff, clinicians in Missouri were significantly more likely to use the single question than the CAGE questions (81% versus 69%, P = 0.001).46 Assuming a prevalence of 20%, this study found that a GP using a single question as the first step, followed by the AUDIT would achieve the same overall accuracy as the AUDIT or CAGE alone but they would require only 121 questions per 100 attendees rather than 1000 with the AUDIT alone and 400 with the CAGE alone (Table 3). Thus an algorithm approach should retain high acceptability, subject to the choice of the first question. In those that screen positive to both steps, consideration should be given to the brief alcohol interventions where required. An updated Cochrane review identified 24 brief alcohol intervention trials for alcohol problems in general practice.48 After ≥1 year, individuals who received a brief intervention drank less alcohol than individuals in the control group (average difference 38 g per week, range 23g–54 g) although the benefit was not clear for females. A cumulative meta-analysis by date of publication demonstrated that effects have shown significance since 1997.57 That said brief alcohol interventions are performed relatively infrequently in primary care.59 Only 10% of GPs reported doing a brief intervention regularly and 50.0% stated they use this approach occasionally. Furthermore, qualitative work on alcohol screening in the primary care setting in Finland by Aira et al. identified that physicians were more comfortable in undertaking a preventive approach for smoking than for alcohol use.59 Factors contributing to this difference were difficulties in recognition and determination of the health risk by physicians, lack of effective tools, and lack of positive feedback after interventions. These must also be considered with any attempts to improve implementation of secondary prevention of alcohol misuse. Even clinicians who participate in successful trials do not necessarily continue with screening and intervention after the study closes.56 It is important to acknowledge that only one study has examined the implementation of single-item screening in primary care. In 126 current drinkers screening rates for alcohol-use disorder increased from 14.6% at baseline to 20.0% after screening implementation and intervention rates rose from 6.3% to 11.8% (P = 0.039).71

Implications for research and practice
From the results of this meta-analysis a cautious recommendation from the authors can be given for one or two verbal question as a screening test for alcohol-use disorder in primary care, but only when paired with a longer screening tool to decide who warrants a brief alcohol intervention. Further research is required to clarify the added value of this approach compared with unaided clinical assessment using a randomised screening implementation study.

---

Funding
None.

Provenance
Freely submitted; externally peer reviewed.

Competing interests
The authors have declared no competing interests.

Discuss this article
Contribute and read comments about this article: www bjgp org letters
REFERENCES


**Appendix 1. Search results OVID (Embase and Medline)**

<table>
<thead>
<tr>
<th>#</th>
<th>Searches</th>
<th>Results</th>
<th>Search Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>alcohol drinking/ or alcohol-induced disorders/ or disorders, nervous system/ or alcohol-related disorders/</td>
<td>90919</td>
<td>Advanced</td>
</tr>
<tr>
<td>2</td>
<td>exp Alcohol/</td>
<td>97021</td>
<td>Advanced</td>
</tr>
<tr>
<td>3</td>
<td>(drink* or alcohol* or substance abuse or substance misuse).m_titl.</td>
<td>179279</td>
<td>Advanced</td>
</tr>
<tr>
<td>4</td>
<td>1 or 2</td>
<td>173857</td>
<td>Advanced</td>
</tr>
<tr>
<td>5</td>
<td>3 and 4</td>
<td>72133</td>
<td>Advanced</td>
</tr>
<tr>
<td>6</td>
<td>limit 5 to ['diagnosis (sensitivity)' or 'diagnosis (specificity)' or 'diagnosis (optimized)']</td>
<td>14267</td>
<td>Advanced</td>
</tr>
<tr>
<td>7</td>
<td>[sensitivity or specificity or ROC or receiver operator or accuracy or youden or valid*].mp. [mp=ti, ot, ab, nm, hw, ui, an, sh, tn, dm, mf]</td>
<td>2249582</td>
<td>Advanced</td>
</tr>
<tr>
<td>8</td>
<td>5 and 7</td>
<td>5767</td>
<td>Advanced</td>
</tr>
<tr>
<td>9</td>
<td>6 or 8</td>
<td>16108</td>
<td>Advanced</td>
</tr>
<tr>
<td>10</td>
<td>(single or question or questions).m_titl.</td>
<td>198813</td>
<td>Advanced</td>
</tr>
<tr>
<td>11</td>
<td>9 and 10</td>
<td>96</td>
<td>Advanced</td>
</tr>
<tr>
<td>12</td>
<td>Primary Health Care/</td>
<td>52388</td>
<td>Advanced</td>
</tr>
<tr>
<td>13</td>
<td>exp Primary Care/</td>
<td>92240</td>
<td>Advanced</td>
</tr>
<tr>
<td>14</td>
<td>[Primary Care or general practit*].mp. [mp=ti, ot, ab, nm, hw, ui, an, sh, tn, dm, mf]</td>
<td>144598</td>
<td>Advanced</td>
</tr>
<tr>
<td>15</td>
<td>12 or 13 or 14</td>
<td>198407</td>
<td>Advanced</td>
</tr>
<tr>
<td>16</td>
<td>11 and 15</td>
<td>14</td>
<td>Advanced</td>
</tr>
<tr>
<td>17</td>
<td>remove duplicates from 16</td>
<td>10</td>
<td>Advanced</td>
</tr>
</tbody>
</table>