**Abstract**

**Purpose:** To review the reliability and validity of the CAGE questionnaire across different patient populations and discuss its role in the detection of alcohol-related problems.

**Methods:** The Cochrane Database for Systematic Reviews, Medline, Embase, and Psychinfo were searched. No systematic reviews were found on the Cochrane Database. Search of the other databases yielded one systematic review and one meta-analysis, on different aspects of CAGE. Three articles on reliability and 16 on validity of CAGE were found and used. Studies generally yielded Level II evidence.

**Results:** CAGE has demonstrated high test-retest reliability (0.80-0.95), and adequate correlations (0.48-0.70) with other screening instruments. The questionnaire is a valid tool for detecting alcohol abuse and dependence in medical and surgical inpatients, ambulatory medical patients, and psychiatric inpatients (average sensitivity 0.71, specificity 0.90). Its performance in primary care patients has been varied, while it has not performed well in white women, prenatal women, and college students. Furthermore, it is not an appropriate screening test for less severe forms of drinking.

**Conclusions:** CAGE is short, feasible to use, and easily applied in clinical practice. However, users should be aware of its limitations when interpreting the results. A positive screen should be followed by a proper diagnostic evaluation using standard clinical criteria.

Alcohol misuse constitutes an important public health problem. Among patients seen by primary care physicians, 10%-36% suffer from alcohol abuse or dependence. Alcohol misuse can lead to social, work-related, or legal problems. In Canada, the alcohol-related death rate in 1995 was 22.2 per 100,000 person-years overall and 4.0 per 100,000 person-years after excluding injuries. In the United States in 2002, the death rate (excluding injuries) was 7.0 per 100,000 person-years.

Detection of alcohol misuse through opportunistic screening is important for prevention of alcohol-related morbidity and mortality. For example, there may be adverse drug-alcohol interactions when alcohol reaches a certain level. Detection of alcohol misuse can also make available information that can be important in clinical research studies, particularly in analysis of effect of behavior and in studies of gene-environment interactions.

In an early stage of alcohol misuse, a simple intervention, such as a brief counseling session delivered by a primary care physician, has proven to be an effective treatment. A number of screening instruments have been used to detect alcohol misuse. The CAGE questionnaire is a brief and popular screening instrument used in clinical practice. Other common instruments include the Alcohol Use Disorders Identification Test (AUDIT), and Michigan Alco-
The CAGE Questionnaire

TABLE 1. The National Institute of Alcohol Abuse and Alcoholism Definitions of Heavy Drinking, Hazardous Drinking, Alcohol Abuse, and Dependence

<table>
<thead>
<tr>
<th>Type of drinking</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy drinking</td>
<td>Men: &gt;14 drinks*/week or &gt;4 drinks/occasion</td>
</tr>
<tr>
<td></td>
<td>Women: &gt;7 drinks*/week or &gt;3 drinks/occasion</td>
</tr>
<tr>
<td>Hazardous drinking</td>
<td>Men: ≥21 drinks/week or ≥7 drinks/occasion at least 3 times/week</td>
</tr>
<tr>
<td></td>
<td>Women: ≥14 drinks/week or ≥5 drinks/occasion at least 3 times/week</td>
</tr>
<tr>
<td>Alcohol abuse</td>
<td>1 or more of the following: 1) failure to fulfill major role obligations at work, school, or home due to recurrent drinking; 2) recurrent drinking in hazardous conditions (e.g., driving a car, operating machinery); 3) recurrent legal problems due to alcohol 4) current use despite recurrent interpersonal or social problems</td>
</tr>
<tr>
<td>Alcohol dependence (Alcoholism)</td>
<td>3 or more of the following: 1) tolerance, withdrawal symptoms, or drinking to relieve withdrawal 2) impaired control 3) drank more or longer than intended 4) increased time spent drinking or recovering 5) continued use despite recurrent psychological or physical problems</td>
</tr>
</tbody>
</table>

Reference

*A standard drink contains 0.6 oz. pure alcohol

Methods

Search strategies

We searched the Cochrane Database for Systematic Reviews, Medline (1966 to present), Embase (1980 to present), and Psychinfo using the following search terms: “CAGE”, “CAGE questionnaire”, “psychiatric status rating scales” and “alcohol”. We retrieved 279 abstracts from Medline, 48 from Embase, and 131 from Psychinfo. Articles were excluded if they were not reliability or validity studies. In addition, articles were included if they met the following inclusion criteria: 1) published in a peer-reviewed journal; 2) written in English; 3) reported reliability or validity measures; 4) used a proper gold standard for validity assessment. Additional articles were retrieved from bibliographic references, and could either be original research articles or review articles.

For original articles on reliability, we included all articles that examined test-retest reliability, and those that examined correlation between CAGE and other instruments. The validity criteria used for the inclusion of the studies in our review were the use of the Diagnostic Interview Schedule (DIS) and the Composite International Diagnostic Interview (CIDI). Less commonly, in some studies that we report, self-report or the use of another screening questionnaire, such as the MAST were used as the criterion standard.

Description of validity measures

The National Institute of Alcohol Abuse and Alcoholism (NIAAA) in the US has developed definitions for several types of problem drinking (Table 1). The Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) and the International Statistical Classification of Diseases, 10th Revision (ICD-10) provide guidelines for diagnosis of alcohol use disorder. The Diagnostic Interview Schedule (DIS) has an alcohol module which includes 20 questions based on the DSM-III-R criteria for alcohol use disorders. The Composite International Diagnostic Interview (CIDI) consists of an alcohol section with a se-
ries of 23 questions based on the DSM-IV criteria. The CIDI can be given in written form, although this method is rarely used. The DSM does not have criteria for less severe forms of drinking and cannot be used as the gold standard for this problem. The instruments described above require large amounts of time and trained personnel for administration.

Content and format of CAGE

The CAGE questionnaire was developed in 1968 by Ewing. The acronym stands for 4 yes/no items constituting the screening test: 1) Have you ever felt that you ought to Cut down on your drinking? 2) Have people Annoyed you by criticizing your drinking? 3) Have you ever felt bad or Guilty about your drinking? 4) Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hangover (Eye-opener)? Individual item responses are scored 0 if the person answers “no” and 1 if the person answers “yes”. The total score can range from 0 to 4. CAGE can be administered in 30 seconds, is easily memorized and non-intimidating. It is often used as an interview but can also be given in written form. However, the dichotomous response format of the CAGE can also be relatively insensitive to small differences in alcohol-related problems. The recommended cutoff for CAGE is ≥2 to screen for alcohol abuse or dependence, although a cutoff of ≥1 has been used in some studies.

Results

No systematic reviews were found on the Cochrane Database. Search of the other databases yielded one systematic review and one meta-analysis on different aspects of CAGE. Apart from these, there were a total of 3 articles on reliability, and 16 on validity of CAGE.

### Cage format

Two studies assessed modifications in the format of CAGE. In one study, a version of CAGE with a general introductory statement (“Please tell me about your drinking”) produced a higher sensitivity than a questionnaire that included a more specific, close-ended introductory question (“How much do you drink”). A second study found no influence of either the wording of the introduction or question sequence on the sensitivity of the instrument.

### Reliability and correlations with other screening tests

Reliability refers to consistency or repeatability of scores and is an indicator of random measurement error. Coefficients > 0.7 or 0.8 for the above are usually regarded as adequate. Test-retest reliability of CAGE (test-retest interval of 7 days) was 0.80 in psychiatric outpatients and 0.95 in a community sample with alcohol use disorders (Table 2). In a U.S. study among clients of a drinking and driving treatment program, the correlations were 0.62 with AUDIT and 0.70 with SMAST (Table 2). Scores on CAGE correlated 0.48 with the AUDIT in a large community sample in the UK.

### TABLE 2. Reliability Studies of the CAGE Questionnaire

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Method</th>
<th>Correlation with AUDIT</th>
<th>Correlation with SMAST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teitelbaum, 2000</td>
<td>Psychiatric sample, 7-day interval</td>
<td>Test-retest correlation</td>
<td>0.62</td>
<td>0.70</td>
</tr>
<tr>
<td>Teitelbaum, 2000</td>
<td>Community sample, 7-day interval</td>
<td>Test-retest correlation</td>
<td>0.80</td>
<td>0.95</td>
</tr>
<tr>
<td>Hays, 1995</td>
<td>Drinking driving treatment program clients, Southern California</td>
<td>Correlation with AUDIT</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Hays, 1995</td>
<td>Drinking driving treatment program clients, Southern California</td>
<td>Correlation with SMAST</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>Hodgson, 2003</td>
<td>Accident and Emergency department patients in 4 U.K. centres</td>
<td>Correlation with AUDIT</td>
<td>0.70</td>
<td>0.48</td>
</tr>
</tbody>
</table>
the concepts measured. Given that these instruments measure slightly different concepts, the observed correlations are in the expected range.

**Validity**

The CAGE was originally validated by Mayfield in a group of 366 psychiatric inpatients (Table 3). Other studies have subsequently assessed the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of CAGE as a screening tool for alcohol abuse/dependence. Positive predictive value is important clinically, as it gives information about those who have disease among those who are screened.

In a meta-analysis of 10 studies, for a cutoff ≥2, the sensitivities were 0.87 in hospital inpatients, 0.71 in primary care patients, and 0.60 in ambulatory medical patients. The specificities were 0.77, 0.91, and 0.92, and the PPVs were 0.57, 0.74, and 0.82, respectively. In Magruder Habib’s study of patients attending a general medical clinic, the PPV was only 0.53 for lifetime alcohol abuse / dependence. Other studies of hospital and ambulatory patients show acceptable values of sensitivity, specificity, and PPV, although these coefficients in primary care patients are varied.

In a large study of college freshmen attending a Catholic University in Belgium, a cutoff ≥1 yielded a sensitivity of only 0.42, and specificity was 0.87, with a PPV of 0.36. College students tend to binge drink and CAGE may not be as useful for detecting this form of alcohol abuse. When the second question of CAGE was replaced with “driving under the influence”, sensitivity increased to 0.94 and specificity was 0.89.

Most studies, although including women, did not examine sex-based differences of CAGE. A study

<table>
<thead>
<tr>
<th>Source</th>
<th>Study Population</th>
<th>Cutoff Score</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Prevalence</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mayfield, 19742, 24</td>
<td>VA hospital psychiatric inpatients</td>
<td>≥2</td>
<td>0.81</td>
<td>0.89</td>
<td>39%</td>
<td>0.82</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥1</td>
<td>0.90</td>
<td>0.72</td>
<td></td>
<td>0.67</td>
<td>0.92</td>
</tr>
<tr>
<td>Bush, 19872, 25</td>
<td>Medical &amp; orthopedic inpatients</td>
<td>≥2</td>
<td>0.75</td>
<td>0.96</td>
<td>20%</td>
<td>0.82</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥1</td>
<td>0.85</td>
<td>0.89</td>
<td></td>
<td>0.62</td>
<td>0.96</td>
</tr>
<tr>
<td>Barry, 199027</td>
<td>Primary care clinics</td>
<td>≥1</td>
<td>0.39</td>
<td>0.93</td>
<td>30%</td>
<td>0.70</td>
<td>0.78</td>
</tr>
<tr>
<td>Beresford, 199028</td>
<td>General hospital medical &amp; surgical patients</td>
<td>≥2</td>
<td>0.76</td>
<td>0.94</td>
<td>27%</td>
<td>0.87</td>
<td>0.91</td>
</tr>
<tr>
<td>Buchsbaum, 19912, 28</td>
<td>Ambulatory medical clinic</td>
<td>≥2</td>
<td>0.74</td>
<td>0.91</td>
<td>36%</td>
<td>0.82</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥1</td>
<td>0.89</td>
<td>0.81</td>
<td></td>
<td>0.72</td>
<td>0.93</td>
</tr>
<tr>
<td>Magruder-Habib, 199330</td>
<td>General medical clinic</td>
<td>Life ≥2</td>
<td>0.78</td>
<td>0.76</td>
<td>25%</td>
<td>0.52</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Present ≥2</td>
<td>1.00</td>
<td>0.61</td>
<td></td>
<td>0.46</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Present ≥3</td>
<td>1.00</td>
<td>0.81</td>
<td></td>
<td>0.64</td>
<td>1.00</td>
</tr>
<tr>
<td>Chan, 199431</td>
<td>Primary care patients</td>
<td>≥2</td>
<td>0.91</td>
<td>0.84</td>
<td></td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>Saitz, 199932</td>
<td>Latinos living in U.S.</td>
<td>≥2</td>
<td>0.80</td>
<td>0.93</td>
<td>36%</td>
<td>0.73</td>
<td>0.89</td>
</tr>
<tr>
<td>Aertgeerts, 200030</td>
<td>College freshmen, Belgium</td>
<td>≥1</td>
<td>0.42</td>
<td>0.87</td>
<td>14%</td>
<td>0.36</td>
<td>0.90</td>
</tr>
<tr>
<td>Fiellin, 20007 (review)</td>
<td>Primary care patients</td>
<td>≥2</td>
<td>0.43-0.94</td>
<td>0.70-0.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saremi, 200133</td>
<td>American Indians</td>
<td>≥2 men</td>
<td>0.68</td>
<td>0.93</td>
<td>85%</td>
<td>0.98</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥2 women</td>
<td>0.62</td>
<td>0.79</td>
<td>53%</td>
<td>0.77</td>
<td>0.65</td>
</tr>
</tbody>
</table>

PPV = Positive Predictive Value
NPV = Negative Predictive Value

**Bold** = Values calculated by authors of this article
among white women in a primary care population yielded a sensitivity of only 0.38. CAGE has tended to perform somewhat better in black females. It has not performed adequately in prenatal women. Sensitivity may be lower in pregnant women and in women in general due to under-reporting, likely due to stigma. Positive predictive values are also lower in women, reflecting a much lower prevalence of alcohol misuse in women. In a study examining male and female drinkers that reported having ≥15 drinks/week, the PPV was 0.46 for men, and only 0.27 for women.

A review by Fiellin showed CAGE to be superior to AUDIT in terms of screening for alcohol abuse/dependence in a primary care population. Only sensitivities and specificities were reported in this review article. CAGE has also been used to detect heavy or hazardous drinking; however, it was less sensitive and specific than AUDIT in head-to-head comparisons.

**Combining CAGE with other questionnaires**

Rumpf et al. reported a study in which 2 CAGE questions were combined with 5 MAST questions to screen for alcohol abuse/dependence in general hospital and general practice samples. Using the alcohol module of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) as the gold standard, this new instrument, called the Luebeck Alcohol Dependence and Abuse Screening test (LAST), demonstrated a higher sensitivity than with the standard CAGE (Table 4).

In a study of 40 year-old men attending health screening in a Finnish town, AUDIT questions 1 and 2 (asking about frequency and quantity of use) were combined with CAGE questions 2, 3, and 4 to yield a new instrument that was used to differentiate between moderate drinkers (<140 g/week) and heavy drinkers (≥280 g/week), as determined from self-reported alcohol consumption. At a cut-off of ≥3, the combined questionnaire gave a sensitivity of 0.77 for detecting heavy drinkers, higher than the CAGE sensitivity of 0.47. Using the same cutoff, the specificity was 0.83 compared to a CAGE specificity of 0.87.

A study of male general medical outpatients used an augmented version of CAGE, in which the 4 CAGE questions and the question “Have you ever had a drinking problem?” were followed by AUDIT ques-

---

**TABLE 4. Validity Studies of Combination Questionnaires**

<table>
<thead>
<tr>
<th>Source</th>
<th>New Questionnaire</th>
<th>Combined * Questionnaire</th>
<th>Study Population</th>
<th>Condition</th>
<th>Cutoff Score</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rumpf, 1997</td>
<td>LAST</td>
<td>CAGE (2) MAST (5)</td>
<td>General hospital</td>
<td>Alcohol Abuse/Dependence</td>
<td>LAST ≥2</td>
<td>0.82</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CAGE ≥2</td>
<td>0.72</td>
<td></td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LAST ≥2</td>
<td>0.63</td>
<td></td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CAGE ≥2</td>
<td>0.53</td>
<td></td>
<td>0.93</td>
</tr>
<tr>
<td>Seppa, 1998</td>
<td>5-Shot AUDIT (2)</td>
<td>CAGE (3)</td>
<td>40 y/o men attending health screening</td>
<td>Heavy Drinking</td>
<td>5-Shot ≥ 3</td>
<td>0.77</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CAGE ≥ 2</td>
<td>0.47</td>
<td></td>
<td>0.87</td>
</tr>
<tr>
<td>Bradley, 1998</td>
<td>Augmented CAGE</td>
<td>CAGE (4) AUDIT (2)</td>
<td>General medical outpatients</td>
<td>Heavy Drinking</td>
<td>CAGE ≥2</td>
<td>0.49</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Augmented-CAGE ≥2</td>
<td>0.70</td>
<td></td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AUDIT ≥8</td>
<td>0.57</td>
<td></td>
<td>0.92</td>
</tr>
</tbody>
</table>

*See appendix
Patients were classified as heavy drinkers if they drank >14 drinks/week in a month, or ≥5 drinks/day at least monthly using the DSM-IIIR as the gold standard. The sensitivities for CAGE, augmented CAGE, and AUDIT were 0.49, 0.70, 0.57, and the specificities were 0.75, 0.68, and 0.92, respectively.1

Discussion

There are a number of different instruments used for screening for alcohol misuse. The CAGE questionnaire has been the most widely used instrument for detecting alcohol abuse / dependence. CAGE has demonstrated high test-retest reliability (0.80-0.95), and adequate correlations with other instruments (0.48-0.70). It also appears to have adequate validity for detecting alcohol abuse/dependence in medical and surgical inpatients, psychiatric inpatients, and ambulatory medical patients. It has not performed well in white women, prenatal women, and college students, and is not recommended as a screening test for heavy or hazardous drinking.

There has been one systematic review8 and one diagnostic meta-analysis9 regarding the CAGE questionnaire, although other review articles have been published.6,10 Other opinions have corroborated the above findings. However, in their overview of screening and diagnosis of alcohol use disorders, Maisto and Saitz state that CAGE ≥1 can be used to detect hazardous drinking10. In practice, however, it may be best to use AUDIT in this situation.6,8,9

Psychometric properties of CAGE in the included studies varied depending on the study population and standards used for validity assessment. Changing the cutoff score from 2 to 1 resulted in greater test sensitivity but lower specificity, as expected. Specificity of a screening test is important, considering the social and legal consequences of false identification of alcohol abuse.33 A cutoff ≥2 is recommended to detect alcohol abuse or dependence to provide the best combination of sensitivity, specificity, and positive predictive value.

Alternative screening questionnaires to CAGE include AUDIT and MAST. AUDIT is currently the only instrument yielding high sensitivities and specificities for less severe forms of drinking. MAST is too long for routine use in clinical practice and more information is needed on the properties of its abbreviated versions (BMAST and SMAST) in different populations.

Limitations

There are a number of methodological problems in the studies included in this review. First, the choice of a gold standard may affect the results of validity studies. While most studies used CIDI or DSM-IIIR for this, some studies used other criteria, for example self-reported alcohol consumption.11 Comparisons between studies that use different validity criteria are limited. Second, comparisons between instruments across studies are difficult to interpret due to methodological differences while head-to-head comparisons are relatively rare. Third, bias due to non-response was a potential problem.1 For example, in one study, non-respondents were significantly more likely to be heavy drinkers and problem drinkers than participants and reported higher scores on the augmented CAGE.1 Fourth, the data may have been influenced by measurement errors due to social desirability, interviewer bias, question misinterpretation, and use of proxy respondents.1,20 Adequate training of the interviewers, and computer-based self-reporting can improve quality of the data and should be implemented in future studies.10 Fifth, generalizability of the results to specific patient groups may be limited. In particular, there was lack of sufficient representation of women, persons <18 years of age, and ethnic minorities. Finally, screening for alcohol misuse in the elderly is not addressed in our article, and this topic would require a separate review.

Future studies

Differences in the psychometric properties of CAGE among different populations need to be further elucidated. It is possible that different screening tests could be used in different groups, for example, based on gender or ethnicity. More studies are needed on combined questionnaires that include CAGE as one
component. Also, the wording of the introduction and the number of response options could be modified to improve the psychometric properties of CAGE.

Conclusions

Although the CAGE questionnaire can be a useful and valid screening tool, many physicians and other health professionals may be reluctant to apply it.9, 37 Our review has established that CAGE has limitations in certain populations and can be used to screen for only certain types of alcohol misuse problems. However, in general, a wider use of this and other brief questionnaires to screen patients for alcohol-related problems in appropriate circumstances would likely improve the provision of care for such patients.

Acknowledgments

The authors would like to acknowledge Dr. Kay Teschke, for comments on an earlier version of the manuscript.

References

3. Canadian Profile: Alcohol, Tobacco, and other Drugs, 1999. Available at: www.ecsa.ca


Correspondence to:
Shayesta Dhalla, MD, MHSc
Department of Health Care and Epidemiology
James Mather Building
5804 Fairview Avenue
Vancouver, B.C., Canada
V6T 1Z3
Tel: 604-822-2772
Fax: 604-822-4994
Email: shayestadhalla@yahoo.com
Appendix

CAGE Questions
1. Have you ever felt you ought to cut down on your drinking?
2. Have people annoyed you by criticizing your drinking?
3. Have you ever felt bad or guilty about your drinking?
4. Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover?

Audit Questions
1. How often did you have a drink containing alcohol in the past year?
   (0) Never
   (1) Monthly or less
   (2) 2-4 times/month
   (3) 2-3 times/week
   (4) ≥4 times/week
2. How many drinks containing alcohol did you have on a typical day when you were drinking in the past year?
   (1) 1 or 2
   (2) 3 or 4
   (3) 5 or 6
   (4) 7 to 9
   (5) 10 or more

Augmented CAGE: 1. CAGE questions
2. AUDIT questions (1-2)
3. Have you ever had a drinking problem?

5-Shot: 1. AUDIT Questions (1-2)
2. CAGE questions (2-4)

**LAST Questionnaire:
1. Are you able to stop drinking when you want to?
2. CAGE Questions (1) and (2)
3. Does any near relative or close friend worry or complain about your drinking?

** All yes or no questions
Question 1 scoring: no
Questions 2-6: scoring: yes
Two or more points indicative of alcohol dependence or abuse.