Validity of Brief Alcohol Screening Tests Among Adolescents: A Comparison of the AUDIT, POSIT, CAGE, and CRAFFT

John R. Knight, Lon Sherritt, Sion Kim Harris, Elizabeth C. Gates, and Grace Chang

Background: Adolescents should be screened for alcohol misuse as part of routine care. The objective of this study was to compare the criterion validity of the Alcohol Use Disorders Identification Test (AUDIT), the Problem Oriented Screening Instrument for Teenagers substance use/abuse scale (POSIT), and the CAGE and CRAFFT questions among adolescents.

Methods: Fourteen- to 18-year-old patients arriving for routine healthcare at a large, hospital-based adolescent clinic completed the four screens and the criterion standard Adolescent Diagnostic Interview, which yields DSM-IV diagnoses of alcohol abuse and dependence. Receiver operating characteristic (ROC) curves were plotted to determine optimal cut-points. Areas under the ROC curves of the four screens were compared, and sensitivities and specificities were calculated.

Results: Participants’ past 12-month alcohol diagnostic classifications were as follows: no use (58.6%), nonproblem use (13.0%), problem use (20.8%), abuse (5.4%), and dependence (2.2%). Optimal cut-points associated with problem use or higher were 2 for AUDIT, 1 for POSIT, 1 for CAGE, and 1 for CRAFFT. ROC curve area of the CAGE was significantly lower compared with areas of all other screens. Sensitivities (95% confidence intervals) were AUDIT 0.88 (0.83–0.93), POSIT 0.84 (0.79–0.90), CAGE 0.37 (0.29–0.44), and CRAFFT 0.92 (0.88–0.96); specificities were AUDIT 0.81 (0.77–0.85), POSIT 0.89 (0.86–0.92), CAGE 0.96 (0.94–0.98), and CRAFFT 0.64 (0.59–0.69).

Conclusions: The AUDIT, POSIT, and CRAFFT have acceptable sensitivity for identifying alcohol problems or disorders in this age group. The CAGE is not recommended for use among adolescents.

Key Words: Substance-Related Disorders, Alcoholism, Substance Abuse Detection, Sensitivity and Specificity, Adolescence.

MISUSE OF ALCOHOL is one of our nation’s greatest health problems with an estimated annual cost of more than $185 billion (Wagenaar et al., 2000) and more than 100,000 alcohol-related deaths each year (National Institute on Alcohol Abuse and Alcoholism, 2001). Heavy drinking adversely affects all ages, including adolescents, and is associated with both serious short- and long-term consequences. Greater than 35% of fatal motor vehicle crashes involving drivers ages 15 to 20 years old are related to alcohol use, and almost 25% involve a legally intoxicated driver (Centers for Disease Control, 1999). Adolescent drinking is linked to other serious health risk behaviors such as early sexual activity, truancy, violence, and weapon carrying (Kokotailo, 1995). Heavy drinking also interferes with normal adolescent cognitive, emotional, and social development and is linked to psychiatric disorders and delinquency (Hicks et al., 1993). Those who begin drinking before age 15 are four times as likely to develop alcohol dependence and more than twice as likely to develop alcohol abuse compared with peers who delay drinking until age 21 (Grant, 1997; Robis and Przybeck, 1985).

Primary care physicians and other healthcare providers can play an important role in preventing the harm associated with alcohol misuse by recognizing problems early and providing appropriate intervention. A number of published guidelines emphasize the importance of routine alcohol screening of adolescents, including the American Medical Association’s Guidelines for Adolescent Preventive Services (Elster and Kuznets, 1994), the Bright Futures Guidelines (Green, 1994), and the American Academy of Pedi-
The Problem Oriented Screening Instrument for Teenagers (POSIT) was developed by the National Institute on Drug Abuse specifically for adolescents aged 12 through 19 (Rahdert, 1991). The POSIT includes a total of 139 yes/no questions divided into 10 scales, one of which is the 17-item substance use/abuse scale (Allen and Columbus, 1995). A score of one or higher on this scale is considered a “red flag” for substance-related problems (Rahdert, 1991). The POSIT has been shown to have good reliability and validity among adolescents, including those seen in primary care medical settings (Dembo et al., 1996; Knight et al., 2001; McLaney et al., 1994).

Although these two questionnaires have advantages, administration in the waiting room may pose a risk to adolescents’ confidentiality (Foster et al., 1997; McLaney et al., 1994). An alternative is an orally administered screen that physicians can administer during the course of the medical visit, after parents have left the room.

The CAGE questions are a good example of this type of brief screen (Ewing, 1984). Originally developed for screening adults, the CAGE has become popular among medical clinicians (Mayfield et al., 1974). Its four yes/no questions are weighted equally (each yes answer = 1), and a total score of two or greater is considered clinically significant (Allen and Columbus, 1995). The CAGE has been well validated among adults and may be particularly suited to identifying patients with alcohol dependence (Bush et al., 1987; Magruder-Habib et al., 1993; Soderstrom et al., 1997). However, the CAGE has been found to perform less well among females, college students, and adolescent medical patients (Aertgeerts et al., 2000; Chung et al., 2000; Knight et al., 2000; O’Hare and Tran, 1997).

The CRAFFT shares many of the CAGE’s advantages but was designed specifically to be developmentally appropriate for adolescents (Knight et al., 1999). CRAFFT is an acronym of the first letters of key words in the test’s six questions: “Have you ever ridden in a CAR driven by someone (including yourself) who was “high” or had been using alcohol or drugs? Do you ever use alcohol or drugs to RELAX, feel better about yourself, or fit in? Do you ever use alcohol or drugs while you are by yourself, ALONE? Do you ever FORGET things you did while using alcohol or drugs? Do your family or FRIENDS ever tell you that you should cut down on your drinking or drug use? Have you ever gotten into TROUBLE while you were using alcohol or drugs?”

Like the POSIT, CRAFFT screens for use of both alcohol and drugs. A previous study concluded that the CRAFFT has adequate sensitivity and specificity for identifying adolescents with substance-related problems (Knight et al., 2002). The purpose of the current study was to compare the criterion validity of the AUDIT, POSIT, CAGE, and CRAFFT in identifying alcohol-related pathology in an adolescent medical clinic population.

METHODS

This criterion standard study was conducted between March 15, 1999, and September 14, 2000, at the Adolescent/Young Adult Medical Practice of Children’s Hospital Boston. This clinic serves both inner city and suburban youth from a wide range of social strata, racial groups, and ethnic backgrounds. Primary care providers, including staff physicians and fellows specializing in adolescent medicine, pediatric residents, and nurse practitioners, served almost 5000 patients through both routine well care and urgent care visits during the study recruitment period.

The study sample was drawn from all 14- to 18-year-old patients arriving for routine care. Before clinic sessions began, the birth dates of all scheduled patients were reviewed and a study recruitment form was placed on the chart cover of each age-eligible patient. The form included items that recorded demographic information, the provider’s impression of the
patient’s level of alcohol use, and the patient’s response to the invitation to participate.

Patients who were unable to read and understand English were excluded, as were those with medical or psychiatric problems that precluded participation in research on the day of the clinic visit. Eligible patients met with a research assistant, who explained the study procedures and obtained signed assent. The Children’s Hospital Boston Committee on Clinical Investigation (institutional review board) waived the requirement for parental consent in accordance with current guidelines for adolescent health research (Department of Health and Human Services, 1991; Society for Adolescent Medicine, 1995).

We identified 711 eligible patients but excluded a total of 41 because of cognitive impairment ($n = 27$), insufficient fluency in English ($n = 9$), severe hearing impairment ($n = 2$), anorexia nervosa ($n = 2$), and psychosis ($n = 1$). Of the remaining 670 who were invited, 538 (80%) agreed to participate. The reasons most commonly cited by the refusers ($n = 132$) included “not enough time” (56%), “not interested” (33%), or “came with a parent” (6%). Participants were told that the purpose of the study was to assess the value of screening questions on alcohol use and that their answers would be kept confidential. They were informed that the research team would notify their provider if a serious problem were identified so that appropriate care could be arranged. After completing the assessment battery, each participant received a $25 merchandise certificate as compensation for his or her time.

The assessment battery included the 10-item AUDIT, the 17-item POSIT substance use/abuse scale (hereinafter referred to as the POSIT), the four CAGE questions, the six CRAFFT questions, and the Adolescent Diagnostic Interview (ADI) (Winters and Henly, 1993). The ADI is a 30- to 90-min structured psychiatric diagnostic interview that yields alcohol- and drug-related diagnoses (i.e., abuse and dependence) according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), the criterion standard for this study (American Psychiatric Association, 1994). The ADI has been shown to have good reliability and validity among adolescents (Winters et al., 1993, 1999). A research assistant first orally administered the CRAFFT and CAGE questions, then monitored participants’ completion of the paper/pencil version of the POSIT and AUDIT, and conducted the ADI interview last. All research assistants were college graduates who underwent a structured initial training and periodic reassessments for competence in ADI administration. Initial training included observation of a model interview, practice sessions until a prescribed level of performance was reached, and a videotaped final assessment that was reviewed by study investigators and the author of the ADI (KW) to ensure competence.

All data were independently entered twice into a specially designed data management program based on Access 97® (Microsoft, Redmond, WA), which included automatic range and logic checks and an entry tracking log. The dual entry files were compared and discrepancies reconciled by checking the original data source. The cleaned dataset was then imported into SPSS v10.0® (SPSS Inc., Chicago, IL) for analysis.

Frequencies of sex, age, race/ethnicity, and provider impression of alcohol involvement in the study sample were compared with those of the refusers group by using $\chi^2$ tests. To preserve adequate cell sizes for analyses, we transformed participant age into a dichotomous variable (i.e., younger youth and older youth) based on the sample median. The provider impression variable was similarly transformed into a trichotomous variable (i.e., no use/nonproblem use, problem use/abuse/dependence, and no impression), because provider impressions of use and dependence were uncommon. The refusers group did not differ significantly from the study sample in distribution of age, sex, race/ethnicity, or provider impressions of alcohol involvement. The study sample was also similar to the entire 14- to 18-year-old clinic population in age and race/ethnicity but included a greater proportion of females (68% vs. 59%, $p < 0.001$).

Participants were classified into five mutually exclusive diagnostic groups based on their pattern of alcohol use during the previous 12 months as recorded in the ADI interview. “No use” was defined as no reported drinking during the past year but no reported alcohol-related problems. “Problem use” was defined as one or more reported alcohol-related problems that did not reach the diagnostic threshold for an alcohol-related disorder. “Abuse” and “dependence” were defined by the corresponding DSM-IV criteria. Each ADI was scored twice, first by study personnel using the standard written instructions and then by computer using an SPSS® syntax algorithm developed by the instrument’s authors (Winters and Henly, 1993). There were no instances of disagreement in alcohol diagnoses.

Frequencies of demographic variables and participants’ diagnostic classifications were computed. Differences in proportions of diagnostic classifications for males versus females and younger versus older adolescents were examined by using Fisher’s exact test. Diagnostic classifications were used to define three alcohol-related screening categories: any problem (i.e., problem use, abuse, or dependence), any diagnosis (i.e., abuse or dependence), and dependence. Sensitivity and specificity were calculated for all three screening categories, and the bootstrap technique was used to estimate 95% confidence intervals (Carpenter and Bithell, 2000; Platt et al., 2000; Rutter, 2000). Confidence intervals could not be computed for the degenerate case, that is, where sensitivity or specificity equaled one.

Receiver operating characteristic (ROC) curves of the AUDIT, POSIT, CAGE, and CRAFFT were plotted. The score associated with the point closest to the upper left corner of the ROC plot, which also indicates the maximum product of sensitivity and specificity, was chosen to indicate the optimal cut-point of each screen. However, this method does not account for the relative cost/benefit ratio of false positives and false negatives. ROC curve areas with asymptotic 95% confidence intervals were calculated as an overall indicator of test accuracy. We are not reporting results of analyses of criterion validity for age, sex, and race/ethnicity subgroups because of inadequate power in this sample.

**RESULTS**

Participants were 68% female, 51% black non-Hispanic, 24% white non-Hispanic, 19% Hispanic, and 7% Asian/other and were distributed almost evenly across years of age (Table 1). Greater than 40% of participants reported drinking during the 12 months preceding the study, 28% had experienced alcohol-associated problems, and almost 8% met criteria for a diagnosis of either alcohol abuse or dependence. Compared with males, more females drank (45% vs. 34%, $p = 0.01$). Compared with younger adolescents, more of the older adolescents drank (52% vs. 31%, $p < 0.001$) and were classified with “problem use” (38% vs. 19%, $p < 0.0001$). Alcohol abuse and dependence diagnoses were not significantly associated with participants’ sex or age.

ROC curve areas, sensitivity, and specificity values for relevant cut-points for all three screening categories (i.e., any problem, any disorder, dependence) are presented in Table 2. The area under the ROC curve reflects the overall accuracy of a screening test.

For the “any problem” screening category, the ROC areas of the AUDIT, POSIT, and CRAFFT were similarly high. The ROC area of the CAGE was significantly lower. The screens (cut-points) with the highest sensitivities were CRAFFT (1), AUDIT (2), and POSIT (1). The CRAFFT “optimal” cut-point was 2 (i.e., highest product of sensitivity × specificity), even though this value yielded significantly lower sensitivity than a score of 1. The sensitivity of the CAGE, even at its lowest possible cut-point (1), was...
significantly lower than that of all other screens. However, the specificity of the CAGE was higher.

For the “any disorder” screening category, the ROC curve areas of the AUDIT, POSIT, and CRAFFT again did not differ significantly from one another. The ROC area of the CAGE was significantly lower than that of the AUDIT and POSIT. The screens (cut-point) with the highest sensitivities were CRAFFT (1), AUDIT (2), and POSIT (1). The sensitivity of the CAGE (1) was lower, but specificity was higher, compared with all other screens. The specificity of the CRAFFT (1) was lower than all other screens, whereas specificity of the POSIT (1) was significantly lower than the CAGE (1) only. Optimal cut-points for “any disorder” were 3 for AUDIT, 2 for POSIT, 1 for CAGE, and 2 for CRAFFT. The sensitivity of the AUDIT was 0.54 and the CAGE 0.37 at their commonly recommended cut-points (8 and 2, respectively).

For the “dependence” category, all screens had equally high sensitivities except for CAGE (2), which was significantly lower than all others. The ROC curve areas of the

---

**Table 1. Demographic Characteristics and Alcohol Diagnostic Classifications of 538 Participants**

<table>
<thead>
<tr>
<th>Category</th>
<th>No Use</th>
<th>Nonproblem use</th>
<th>Problem use</th>
<th>Abuse</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (n = 538)</td>
<td>315 (58.6)</td>
<td>70 (13.0)</td>
<td>112 (20.8)</td>
<td>29 (5.4)</td>
<td>12 (2.2)</td>
</tr>
<tr>
<td>Sex Male (n = 170)</td>
<td>113 (66.5)</td>
<td>17 (10.0)</td>
<td>30 (17.6)</td>
<td>7 (4.1)</td>
<td>3 (1.8)</td>
</tr>
<tr>
<td>Female (n = 368)</td>
<td>202 (54.9)</td>
<td>53 (14.4)</td>
<td>82 (22.3)</td>
<td>22 (6.0)</td>
<td>9 (2.4)</td>
</tr>
<tr>
<td>Age 14 Years (n = 103)</td>
<td>85 (82.5)</td>
<td>9 (8.7)</td>
<td>4 (3.9)</td>
<td>4 (3.9)</td>
<td>1 (1.0)</td>
</tr>
<tr>
<td>15 Years (n = 85)</td>
<td>56 (65.9)</td>
<td>14 (16.5)</td>
<td>10 (11.8)</td>
<td>4 (4.7)</td>
<td>1 (1.2)</td>
</tr>
<tr>
<td>16 Years (n = 121)</td>
<td>68 (56.2)</td>
<td>15 (12.4)</td>
<td>28 (23.1)</td>
<td>8 (6.6)</td>
<td>2 (1.7)</td>
</tr>
<tr>
<td>17 Years (n = 127)</td>
<td>61 (48.0)</td>
<td>20 (15.7)</td>
<td>35 (27.6)</td>
<td>7 (5.5)</td>
<td>4 (3.1)</td>
</tr>
<tr>
<td>18 Years (n = 102)</td>
<td>45 (44.1)</td>
<td>12 (11.8)</td>
<td>35 (34.3)</td>
<td>6 (5.9)</td>
<td>4 (3.9)</td>
</tr>
<tr>
<td>Race/ethnicity Black non-Hispanic (n = 272)</td>
<td>180 (66.2)</td>
<td>29 (10.7)</td>
<td>51 (18.8)</td>
<td>7 (2.6)</td>
<td>5 (1.8)</td>
</tr>
<tr>
<td>White non-Hispanic (n = 130)</td>
<td>56 (43.1)</td>
<td>21 (16.2)</td>
<td>37 (28.5)</td>
<td>12 (9.2)</td>
<td>4 (3.1)</td>
</tr>
<tr>
<td>Hispanic (n = 101)</td>
<td>60 (59.4)</td>
<td>11 (10.9)</td>
<td>19 (18.8)</td>
<td>8 (7.9)</td>
<td>3 (3.0)</td>
</tr>
<tr>
<td>Asian/other (n = 35)</td>
<td>19 (54.3)</td>
<td>9 (25.7)</td>
<td>5 (14.3)</td>
<td>2 (5.7)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

Data presented as n (%); all diagnostic classifications based on Adolescent Diagnostic Interview.

---

**Table 2. ROC Curve Area, Sensitivity, and Specificity With 95% Confidence Intervals (95% CI) of Alcohol Screening Test Cut-Points for Three Screening Categories: Any Problem, Any Disorder, and Dependence**

<table>
<thead>
<tr>
<th>Test</th>
<th>Total n (%)</th>
<th>Any problem*</th>
<th>Any disorder*</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUDIT</td>
<td>538 (100)</td>
<td>.92 (.89, .94)</td>
<td>.91 (.87, .95)</td>
<td>.95 (.91, .99)</td>
</tr>
<tr>
<td>POSIT</td>
<td>538 (100)</td>
<td>.88 (.85, .92)</td>
<td>.93 (.89, .96)</td>
<td>.95 (.91, .98)</td>
</tr>
<tr>
<td>CAGE</td>
<td>538 (100)</td>
<td>.67 (.61, .72)</td>
<td>.77 (.67, .86)</td>
<td>.87 (.74, .99)</td>
</tr>
<tr>
<td>CRAFFT</td>
<td>538 (100)</td>
<td>.88 (.85, .92)</td>
<td>.88 (.83, .93)</td>
<td>.89 (.83, .95)</td>
</tr>
<tr>
<td>AUDIT (2)</td>
<td>153 (28)</td>
<td>.88 (.83, .93)</td>
<td>.93 (.84, 1.0)</td>
<td>1.0 (***)</td>
</tr>
<tr>
<td>AUDIT (3)</td>
<td>153 (28)</td>
<td>.72 (.65, .79)</td>
<td>.88 (.76, .97)</td>
<td>1.0 (**)</td>
</tr>
<tr>
<td>AUDIT (5)</td>
<td>153 (28)</td>
<td>.50 (.43, .58)</td>
<td>.73 (.58, .87)</td>
<td>.83 (.57, 1.0)</td>
</tr>
<tr>
<td>AUDIT (8)</td>
<td>153 (28)</td>
<td>.24 (.18, .31)</td>
<td>.54 (.38, .69)</td>
<td>.75 (.46, 1.0)</td>
</tr>
<tr>
<td>POSIT (1)</td>
<td>538 (100)</td>
<td>.84 (.79, .90)</td>
<td>.98 (.92, 1.0)</td>
<td>1.0 (**)</td>
</tr>
<tr>
<td>POSIT (2)</td>
<td>538 (100)</td>
<td>.65 (.58, .72)</td>
<td>.88 (.76, .97)</td>
<td>1.0 (**)</td>
</tr>
<tr>
<td>POSIT (3)</td>
<td>538 (100)</td>
<td>.48 (.40, .56)</td>
<td>.80 (.67, .91)</td>
<td>.92 (.71, 1.0)</td>
</tr>
<tr>
<td>CAGE (1)</td>
<td>538 (100)</td>
<td>.37 (.29, .44)</td>
<td>.61 (.45, .76)</td>
<td>.83 (.57, 1.0)</td>
</tr>
<tr>
<td>CAGE (2)</td>
<td>538 (100)</td>
<td>.18 (.12, .24)</td>
<td>.37 (.22, .50)</td>
<td>.42 (.14, .70)</td>
</tr>
<tr>
<td>CRAFFT (1)</td>
<td>538 (100)</td>
<td>.92 (.88, .96)</td>
<td>.98 (.92, 1.0)</td>
<td>1.0 (**)</td>
</tr>
<tr>
<td>CRAFFT (2)</td>
<td>538 (100)</td>
<td>.70 (.63, .77)</td>
<td>.83 (.71, .94)</td>
<td>.92 (.71, 1.0)</td>
</tr>
<tr>
<td>AUDIT (2)</td>
<td>538 (100)</td>
<td>.81 (.77, .85)</td>
<td>.66 (.62, .70)</td>
<td>.63 (.59, .67)</td>
</tr>
<tr>
<td>AUDIT (3)</td>
<td>538 (100)</td>
<td>.89 (.86, .92)</td>
<td>.77 (.73, .80)</td>
<td>.73 (.70, .77)</td>
</tr>
<tr>
<td>AUDIT (5)</td>
<td>538 (100)</td>
<td>.97 (.95, .99)</td>
<td>.88 (.85, .91)</td>
<td>.85 (.82, .88)</td>
</tr>
<tr>
<td>AUDIT (8)</td>
<td>538 (100)</td>
<td>1.0 (99.9, 1.0)</td>
<td>.97 (.95, .98)</td>
<td>.94 (.92, .98)</td>
</tr>
<tr>
<td>POSIT (1)</td>
<td>538 (100)</td>
<td>.89 (.86, .92)</td>
<td>.73 (.69, .77)</td>
<td>.69 (.65, .74)</td>
</tr>
<tr>
<td>POSIT (2)</td>
<td>538 (100)</td>
<td>.94 (.91, .96)</td>
<td>.82 (.79, .86)</td>
<td>.79 (.75, .82)</td>
</tr>
<tr>
<td>POSIT (3)</td>
<td>538 (100)</td>
<td>.97 (.95, .99)</td>
<td>.90 (.87, .92)</td>
<td>.86 (.83, .89)</td>
</tr>
<tr>
<td>CAGE (1)</td>
<td>538 (100)</td>
<td>.96 (.94, .98)</td>
<td>.91 (.88, .93)</td>
<td>.88 (.86, .91)</td>
</tr>
<tr>
<td>CAGE (2)</td>
<td>538 (100)</td>
<td>.99 (.98, 1.0)</td>
<td>.97 (.95, .98)</td>
<td>.95 (.93, .97)</td>
</tr>
<tr>
<td>CRAFFT (1)</td>
<td>538 (100)</td>
<td>.64 (.59, .69)</td>
<td>.52 (.47, .56)</td>
<td>.49 (.45, .53)</td>
</tr>
<tr>
<td>CRAFFT (2)</td>
<td>538 (100)</td>
<td>.94 (.91, .96)</td>
<td>.81 (.77, .84)</td>
<td>.77 (.74, .81)</td>
</tr>
</tbody>
</table>

* Any problem = alcohol problem use, abuse or dependence; any disorder = alcohol abuse or dependence.
** 95% CI cannot be computed.
four screens did not differ significantly when screening for dependence.

**DISCUSSION**

This study provides strong supportive evidence for the criterion validity of the AUDIT, POSIT, and CRAFFT questions in an adolescent general outpatient clinic sample. Providers should carefully consider test characteristics, the level of alcohol-related pathology they wish to identify, and practical limitations of test administration and scoring before deciding which screen best suits the needs of their particular clinical practice. Sensitivity may be the single most important characteristic to consider when screening for alcohol disorders, because the risks of a false negative (i.e., missed diagnosis) far outweigh those of a false positive (i.e., additional assessment), although some adolescents may find additional assessment distressing. The CAGE does not have sufficient sensitivity and its use among adolescents is not recommended.

Our findings also indicate that commonly used cut-points for adults must be lowered when alcohol screens are administered to adolescents. The AUDIT recommended cut-point of 8 was found to be too high. Studies in college populations have recommended cut points of 6 and 11 (Aertgeerts et al., 2000; Fleming et al., 1991), and Chung et al. (2000) reported 4 as optimal in a study of 13- to 19-year-old emergency ward patients. We found that a cut-point of 2 was optimal for identifying alcohol problem use and 3 for identifying alcohol abuse or dependence. Compared with Chung’s study, our sample had a narrower age range and a lower prevalence of alcohol disorders (8% vs. 18%), which may account for variations in sensitivity. Similarly, the standard recommendation for the CAGE is a cut-point of 2. We found the sensitivity of CAGE too low even when its cut-point is lowered to 1.

Test characteristics, including sensitivity, vary across screening categories, so providers should carefully consider the purpose of their screening before selecting a particular test. Do they wish to identify adolescents with alcohol problems, alcohol disorders, or alcohol dependence only? In most circumstances, providers will likely choose to screen for all alcohol problems and disorders (i.e., the “any problem” screening category in this study). Practical aspects of test administration and scoring are also important considerations. Each of the three recommended screening tests presents its own unique advantages and disadvantages. The AUDIT, for example, is the only recommended tool that screens for alcohol only. This may be desirable in certain situations, for example, when clinicians wish to identify adolescents who are possible candidates for a brief alcohol intervention. In most circumstances, however, the limitation of screening for alcohol only presents a disadvantage. Alcohol use is highly associated with illicit drug use, and therefore screening for all substances simultaneously seems most practical. The AUDIT also requires staff time for scoring, although a computerized version could easily overcome this limitation. The AUDIT may offer an advantage in efficiency in clinic settings where patients can self-administer the test while waiting to see the provider. However, clinics must designate a private place for this purpose to ensure complete confidentiality.

The POSIT is another written or computer-based questionnaire that possesses advantages and limitations similar to those of the AUDIT. Unlike the AUDIT, however, the POSIT screens for alcohol and drugs simultaneously. The full 139-item version of the POSIT can provide a wealth of additional, clinically relevant information from its nine other scales, but it requires 20 to 30 min for completion. The written version of the POSIT requires a substantial commitment of staff resources for scoring, whereas the computerized version does not. In fact, the computerized POSIT (PowerTrain, Inc., Landover, MD) automatically prepares a convenient, summary bar graph that compares individual scale scores with normative cut-points. This summary may be printed out and used to initiate a discussion between the adolescent and provider. The POSIT substance use/abuse scale’s recommended cut-point of 1 (Rahdert, 1991) yielded high sensitivity and specificity in our study, although a score of 2 yielded the highest product of these two characteristics. This is identical to the optimal cut-point (2) reported by Latimer et al. (1996) in a sample of adolescents from schools, clinics, and correctional settings. Latimer’s report also found similar sensitivity (0.95) and specificity (0.79) for POSIT in identifying substance use disorders to those found in this study. The POSIT is a valid screening tool for adolescents at either cut-point.

The CRAFFT screen also presents unique advantages. Like the CAGE, it has a convenient mnemonic and requires only 1 or 2 min for administration and scoring. Like the POSIT, it screens for alcohol and drugs simultaneously. The CRAFFT is brief, and many providers will find it possible to ask the six questions during the course of the medical visit, sparing other staff’s time. A CRAFFT score of 1 demonstrates high sensitivity for identifying all three screening categories. The CRAFFT yielded lower specificity compared with the other screens. Providers should therefore follow positive responses to individual CRAFFT items with additional questions and decide what level of intervention, if any, their adolescent patient requires based on all the information. Raising the CRAFFT score cut-point to 2 provides a better balance between sensitivity and specificity, especially for identifying alcohol disorders.

This study presents certain strengths. It was conducted in a primary care setting, where routine screening is required. The study sample included patients arriving for both routine care and urgent care visits. The criterion standard was a well-validated diagnostic interview for adolescents, so that all screens could be compared directly with DSM-IV diagnostic criteria. All screens were administered before the diagnostic interview, and there were no preliminary questions on alcohol quantity and frequency, which might
lower subsequent screening test scores (Steinweg, 1993). In addition to abuse and dependence disorders, our analysis included a “problem use” diagnostic classification, which has been previously suggested by the American Academy of Pediatrics (Wolraich et al., 1996).

Several limitations should also be noted. The study was conducted in one clinic only, and prevalence rates of alcohol problems and disorders in other clinic settings may be different. Our sample included a large proportion of racial and ethnic minorities, and females coming for medical care may be at higher risk than those in the general population. However, we would expect to find similar sensitivity and specificity for all tests in similar urban medical outpatient clinic populations. Our study results cannot be generalized to early adolescents (10–13 years of age) because they were not included in the sample and screening test characteristics may differ significantly in a younger population. Future studies should address these limitations by including a variety of clinical settings and adolescents less than 14 years of age. We instructed providers to invite all their patients to participate. However, we cannot assess the degree to which they followed this instruction, and selection bias remains a possibility. Our study may also be limited in that it relied on adolescents’ self-report. The extent to which some participants may have overstated and others understated their use of alcohol is unknown. However, previous studies have shown that self-report compares favorably with other methods of substance use detection (Babor et al., 1989; Winters et al., 1990). Obtaining laboratory testing for alcohol was impractical in this study, and obtaining collateral reports of alcohol use from family members poses too great a risk to adolescents’ confidentiality. In addition, self-report may be influenced by order of administration. Since we administered the screens to all participants in an identical sequence, we cannot assess the nature or degree of possible bias from ordering.

In certain aspects, this study is further limited by the size of the study population. Our original sample size calculation was based on achieving adequate power to determine with 0.95 confidence that screening test sensitivities were at least 0.80 across the entire study sample. This model is relatively underpowered for examining differences among demographic subgroups. Future studies should also address this limitation and place special importance on examining possible gender differences in screening test characteristics. Females account for approximately two thirds of patients and three fourths of visits at our clinic, and other practices are similar (Cherry et al., 2001). Brief alcohol screening tests must therefore demonstrate high sensitivities among females, because they use health services more frequently than males.

We conclude that the AUDIT, POSIT, and CRAFFT have good criterion validity for screening adolescent medical patients for alcohol problems and disorders. We also conclude that the CAGE questions are not appropriate for screening adolescents. The CRAFFT questions present a practical and valid alternative to CAGE.

ACKNOWLEDGMENTS

We thank Erin Gates, BA, Sarah Rosenberg, BA, Katherine O’Connor, BA, and Allison Arneill, MA, for their assistance in study implementation; the clinicians and staff of the Adolescent/Young Adult Medical Practice at Children’s Hospital Boston for assistance in recruitment; Ken C. Winters, PhD, for consultation on the study measurement battery, and S. Jean Emans, MD, for review of the manuscript.

REFERENCES


