Reliabilities of Short Substance Abuse Screening Tests Among Adolescent Medical Patients

John R. Knight, MD*; Elizabeth Goodman, MD‡; Todd Pulerwitz, MD§; and Robert H. DuRant, PhD||

Abstract. Objective. To determine the internal consistency and 1-week test-retest reliability of the Simple Screening Instrument for Alcohol and Other Drug Abuse (SSI-AOD), the CAGE-AA (CAGE questions adapted for adolescents), and 4 modified items from the Drug and Alcohol Problem QuickScreen (DAP-4) among adolescents.

Methods. Fifteen- to 18-year-old medical patients (n = 173) completed screening tests during a routine medical visit and then again 1 week later. Internal consistency for each test and retest was calculated using Cronbach’s α, and 1-week test-retest reliability was calculated by using Winer’s unbiased estimate of the intraclass correlation coefficient (r).

Results. The SSI-AOD has good internal consistency (α = .83) and the CAGE-AA questions acceptable internal consistency (α = .60). Alpha varied with gender and race, and item analysis indicated the CAGE-AA test could be improved. As expected, the DAP-4 had a lower α score (.46). All screening instruments studied had high 1-week test-retest reliabilities (range r = .82–.90).

Conclusions. The SSI-AOD is a reliable substance abuse screening instrument among adolescent medical patients. The CAGE-AA questions must be further revised and tested before their use can be recommended. The DAP-4 questions are likely measuring different, but important, constructs.

Abbreviations. AOD, alcohol and other drug (use); TIP, treatment improvement protocols; SSI-AOD, Simple Screening Instrument for Alcohol and Other Drug Abuse; DAP, Drug and Alcohol Problem (QuickScreen); POSIT, Problem Oriented Screening Instrument for Teenagers.

Alcohol and other drug (AOD) use are associated with serious morbidity and mortality among young people. Seventy-three percent of deaths among youth in the United States are the result of only 4 causes: motor vehicle crashes, other unintentional injuries, homicides, and suicides.1 Many of these deaths are related to the use of AOD. According to the 1997 Youth Risk Behavior Survey, 50.8% of the young people surveyed drank alcohol during the preceding 30 days, 26.2% smoked marijuana, and 36.6% rode in a car with a driver who had been drinking.1 This is particularly significant, as >40% of motor vehicle accident deaths in the high school age group (and among adults) are associated with alcohol use.2,3

According to the American Medical Association’s Guidelines for Adolescent Preventive Services (GAPS), Bright Futures, and other guidelines, every adolescent should be screened for use of AOD as part of routine medical care and given appropriate counseling.4,5 In fact, the American Academy of Pediatrics’ policy statement on substance abuse indicates that all medical care providers should be able to determine the degree of risk, offer brief advice, and refer adolescents who are in need to appropriate substance abuse treatment.6 Providers need an efficient and reliable means for accomplishing this. A number of instruments are available for screening adolescents,7 but relatively little research has been conducted to determine validity and reliability of these devices in general adolescent populations.

The Quality Assurance and Evaluation Branch of the Center for Substance Abuse Treatment publishes a series of Treatment Improvement Protocols (TIPs) “to facilitate the transfer of state-of-the-art protocols and guidelines for the treatment of AOD abuse from acknowledged clinical, research, and administrative experts to the nation’s AOD abuse treatment resources.” TIP 11 contains a recommended 16-item screening instrument for AOD abuse.8 This Simple Screening Instrument for Alcohol and Other Drug Abuse (SSI-AOD) is intended for use by a wide variety of service providers, including physicians and nurses, in a broad range of at-risk populations. The SSI-AOD has not been validated, however, and its reliabilities are unknown.

Another test, known as the CAGE, is very popular among adult serving medical care providers as a method of screening for alcohol problems.9,10 This instrument’s name is a mnemonic of the following four yes/no questions:

“Have you ever felt that you should CUT DOWN on your drinking?”
“Have people ANNOYED you by criticizing your drinking?”

“Have you ever felt bad or GUILTY about your drinking?”

“Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover (EYEOPENNER)?”

CAGE is brief, verbally administered, easily remembered, and simple to score (each yes answer = 1). CAGE has been shown to have adequate sensitivity and specificity among adult medical patients, and a score of 2 or greater has been shown to predict a high likelihood of an alcohol-related diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised (DSM-III).

CAGE has been evaluated among college freshmen and found to adequately identify male students with more severe alcohol problems. Although it was found to be a useful part of a composite scale among female students, however, Werner and colleagues concluded that CAGE should probably not be used alone as a screening test among older adolescents. CAGE has not been evaluated among early or middle adolescents and some items (eg, the Eye-opener question) are, in fact, not developmentally appropriate for use in this age group. The test is further limited as a substance abuse screening instrument as its items are worded to inquire about alcohol use only.

Another questionnaire, the Drug and Alcohol Problem (DAP) QuickScreen, consists of 30 yes/no items and was developed for use in primary care medical offices. Its questions are worded to inquire about both drug and alcohol use. The DAP was tested among middle-upper socioeconomic-class adolescents in a suburban private pediatric office. Schwartz and Wirtz found that 4 DAP items accounted for 70% of the variation between high-risk and low-risk users:

“Do you use tobacco products (cigarettes, snuff, etc)?”

“Have you ever had an in-school or out-of-school suspension for any reason?”

“Do you sometimes ride in a car driven by someone (including yourself) who is high or who appears to have had too much to drink?”

“Has anyone (friend, parent, teacher or counselor) ever told you that they believe that you may have a drinking or drug problem?”

These 4 items, however, have not been tested independently from the parent instrument.

The purpose of this study was to determine the internal consistency and 1-week test-retest reliability of the SSI-AOD, adapted CAGE questions (CAGE-AA), and 4 modified DAP items (DAP-4) in a general adolescent medical clinic population. Internal consistency indicates that items within a scale are measuring the same construct and that a higher total score is likely to indicate higher total risk. Test-retest reliability measures the temporal stability of a scale, and high reliability indicates that the measurement error of the test is relatively small over brief intervals of time during which behavior itself is unlikely to have changed. Overall, the results of this study will assist clinicians and researchers in determining the utility of these various screening tests in settings where adolescents receive routine care.

METHODS

Subjects

The subjects (n = 173) were 15- to 18-year-old patients receiving medical care in the Adolescent/Young Adult Medical Practice at Children’s Hospital in Boston during June through August 1995. This clinic serves both inner-city and suburban youth from the full range of the social strata and has >4,000 patients and 11,000 visits per year.

Questionnaire

The entire study questionnaire included the 139-item Problem Oriented Screening Instrument for Teenagers (POSIT), the 16-item SSI-AOD, 4 adapted CAGE questions, 4 modified DAP items, 9 sexual risk questions and a 13-item life optimism test. (We have reported results of the other tests elsewhere, and they will not be discussed here). Two questions from the SSI-AOD closely resembled items from the CAGE (“Have you ever tried to Cut down or quit drinking or using drugs? Do you feel bad or Guilty about your drinking/drug use?”). To avoid redundancy, these 2 items were considered part of both the SSI-AOD and the CAGE, and the 2 other CAGE questions were adapted so that they were similarly worded before being added to the questionnaire (“Have you ever been Annoyed with someone because they criticized your drinking or use of drugs?” “Do you ever refer alcohol or drugs Early in the day?”). The result was a significantly modified CAGE test, adapted for adolescents (CAGE-AA). The 4 items from the DAP were also added to the study questionnaire, although 2 were modified to provide consistency among groups of items (ie, “Have you used tobacco products? Have you ever ridden in a car driven by someone who was high or appeared to have been using alcohol or another drug?”) All items on the questionnaire required a yes/no response, were equally weighted and (except for the POSIT) scored in the same direction (yes response = 1 point). The SSI-AOD, CAGE-AA and DAP-4 questions were grouped together in one section of the questionnaire, but items from each were interspersed with each other to enhance overall flow. The entire questionnaire (including POSIT) required approximately 30 minutes for completion. It was pilot-tested among a small group of older college students.

Procedures

A research associate who was not involved in providing medical care consecutively invited patients who were being seen for well visits and general medical problems to participate in the study. The research associate obtained informed consent from the adolescent, as the Children’s Hospital Committee on Clinical Investigations (institutional review board equivalent) waived the requirement for parental consent based on the published Guidelines for Adolescent Health Research. He explained to each prospective subject that the purpose of the study was to measure the reliability of the screening tests, and that he/she would fill out the questionnaire anonymously at the present clinical visit and then again 1 week later. The research associate explained that parents would not be given any information about specific responses to questions unless we found out that someone was in danger. However, he encouraged each subject to tell his/her parents or another responsible adult about participating in the study. At the conclusion of each study visit, the research associate offered each subject a confidential referral to a clinic provider to discuss AOD use or any other issues that were raised during completion of the study questionnaire.

The research associate did not invite patients to participate in the study if their medical care provider judged they would be unusually stressed on the day of the present visit. On this basis, we excluded 3 patients with acute anxiety or depression, 1 patient who came to be examined for alleged sexual assault and 3 others who came to discuss results of a positive pregnancy test. We did not collect data on the number of patients who were invited to participate but refused. However, the research associate esti-
mated that the percentage of refusers was very low after the first 2 weeks of the study (when he was learning how best to approach subjects) and that those who refused typically cited lack of time on the day of the clinic visit as the reason. Potential subjects were offered a small ($5 average value) gift certificate for a local fast-food restaurant as an incentive for participating. The research associate distributed these to subjects after completion of the retest.

Each subject completed the questionnaire at a private desk in the clinic, not in the waiting room. The research associate remained available at all times to explain individual items and answer questions. No subjects, including those from Spanish-speaking families, reported difficulty in understanding items with the exception of 1 young man with mild mental retardation. We did not assess acceptability of the questionnaire, but the research associate did not report any negative comments. We assigned each subject a unique numerical identification number so that we could link test and retest data entries but still protect his/her confidentiality. After the completion of the questionnaire, each subject was given an appointment for the retest 1 week later. The research associate asked for permission to telephone him/her the day before this appointment as a reminder and he tried to reschedule the retest within 8 days of the initial test when a subject did not keep the appointment. The research associate followed the same procedure in administering the retest questionnaire as that used in administering the initial test.

The study investigators performed all statistical computations, except test-retest reliability, using Statistical Package for Social Scientists (SPSS, Chicago, IL) for Windows statistical software. We calculated the frequency of responses for demographic variables and each individual item on the questionnaire at test and retest, and the mean and standard deviation for age, grade in school, and total score of each AOD screening test. We excluded from each analysis (for that scale only) subjects who answered <75% of the questions in that scale. We computed the distribution of subjects’ total scores on each of the screening tests, and the distribution of subjects by SSI-AOD cut-points for low risk (score = 0–1), minimal risk (2–3), and moderate to high risk (≥4). To measure internal consistency, we calculated Cronbach’s α and α-if-item-deleted for all scales and for important demographic subgroups. Finally, we calculated test/retest reliability using Winer’s unbiased estimate of the intraclass correlation coefficient (r). This is the best measure of test-retest reliability of a scale. We considered using Kappa for measuring agreement beyond chance of individual items, but it is affected by the prevalence of the response and cannot be computed when a zero cell occurs in the agreement diagonal (both were considered in the present study).22

RESULTS

Subjects (n = 173) were 71% female, 43% black, 38% Hispanic, and 16% white (Table 1). The Hispanic group consisted of both white-Hispanic and black-Hispanic subjects. These frequencies reflect the demographic distribution of our clinic’s visits, but whites were relatively underrepresented and Hispanics overrepresented in this study sample compared with the clinic population at-large. This discrepancy may have resulted in part to subjects’ confusion over questionnaire structure, ie, there were separate items asking about race (white, black, Asian, etc.) and ethnicity (Hispanic, non-Hispanic). Demographic data, and the means and standard deviations of each AOD screening test score, for the total sample (n = 173) and for the subjects who completed the retest (n = 93) are shown in Table 1. Compliance with the retest (54%) was low and likely attributable to the fact that it required a special return trip to the clinic when none would otherwise have been necessary. The subgroup of those who completed the study differed somewhat from the total sample in gender and race. Females, white, and black subjects were more likely than males and Hispanic subjects to return for the retest. More importantly, however, the retest group did not differ substantially from the total sample on AOD screening test mean scores, suggesting that sample bias in calculation of test-retest reliabilities is minimal.

Distribution of responses for questions that were most frequently answered as positive on the initial test are shown in Table 2, along with data on 2 items that are especially concerning; 6% percent of the study sample (n = 11) had experienced a serious problem as a result of AOD use (blackout, injury, emergency department visit, arrest, etc) yet only 2% (n = 3) had gone to anyone for help. Of those subjects who reported a serious problem, a greater proportion (3 of 11) had gone for help. This is relatively encouraging, but still indicates that a sizeable majority (72.7%) of our patients with recent and serious AOD-related problems have not reached out for help.

Table 3 lists the frequency of total scores for each of the AOD screening tests. According to suggested SSI-AOD cutpoints, 61.8% of our subjects were classified as low risk (score = 0–1), 20.1% as minimal risk (score = 2–3), and 18.1% as moderate to high risk (score ≥4). By way of comparison, 70.5% of subjects had no positive answers to the CAGE-AA questions, while 17.1% had 1 positive

| TABLE 1. | Comparison of Test and Retest Group Demographics and Screening-Test Score Means |
| --- | --- | --- | --- | --- |
| | Test | Retest |
| Demographic variable | | | |
| Age (mean ± SD) | 16.3 ± 1.1 | 16.2 ± 1.0 |
| Grade (mean ± SD) | 10.7 ± 1.4 | 10.6 ± 2.1 |
| Female (%) | 70.8 | 75 |
| Black (%) | 42.8 | 48.4 |
| White (%) | 16.2 | 23.7 |
| Hispanic (%) | 38.2 | 26.9 |
| Asian/Pacific Islander (%) | .6 | 1.1 |
| Indian/Alaskan (%) | 2.3 | 0 |

<table>
<thead>
<tr>
<th>Screening Test</th>
<th>No. of Items</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSI-AOD</td>
<td>16</td>
<td>1.8</td>
<td>2.1</td>
<td>144</td>
<td>1.3</td>
<td>2.2</td>
<td>82</td>
</tr>
<tr>
<td>CAGE-AA</td>
<td>4</td>
<td>1.5</td>
<td>.9</td>
<td>146</td>
<td>.4</td>
<td>.8</td>
<td>83</td>
</tr>
<tr>
<td>DAP-4</td>
<td>4</td>
<td>.9</td>
<td>1.0</td>
<td>160</td>
<td>.8</td>
<td>.9</td>
<td>90</td>
</tr>
</tbody>
</table>
answer and 12.3% had 2 or more positive answers. Despite the fact that mean scores were similar on test and retest (above), frequency data indicate that a greater percentage of subjects with SSI-AOD scores of 0–1 were compliant with the retest (test = 61.6% vs retest = 76.8%).

The internal consistency of each scale was computed using Cronbach’s $\alpha$ (Table 2). The SSI-AOD had good internal consistency among all subjects and among all demographic subgroups, although the $\alpha$ score was relatively lower for males and Hispanics. The $\alpha$ for this scale increased substantially in the smaller number of subjects who completed the retest, suggesting that compliant subjects may answer questions more consistently over time. Item analysis for the total study sample and each demographic subgroup did not suggest that the scale should be refined ($\alpha$-if-item-deleted range = .72–.82). The CAGE- AA questions had acceptable reliability at both test and retest for all subjects ($\alpha$ = .60 and .63, respectively), but an unacceptable $\alpha$ for males ($\alpha$ = .42) and Hispanics ($\alpha$ = .43). Alpha is in part a function of scale length, however, so this finding is not very surprising. The results of the item analysis suggest that internal consistency could be improved by eliminating the first question (“Have you tried to cut down or quit drinking or using drugs?”) for males ($\alpha$ = .53) and the third question (“Do you feel bad or guilty about your drinking or drug use?”) for Hispanics ($\alpha$ = .58). The 4 questions taken from the DAP were never intended to be used independently in a scale, so it is not surprising that their internal consistency was in the unacceptable range for all subgroups but whites ($\alpha$ = .68) and for the entire group of subjects at both test and retest (Table 4).

Test-retest reliability was computed using Win- ner’s unbiased estimate of the intraclass correlation coefficient formula 9.21 High test-retest reliabilities were found for all screening tests, with a range from $r = .82$ to $r = .90$ (Table 2). The highest intraclass correlations were found for the SSI-AOD and DAP-4, and the lowest for the CAGE- AA questions.

**DISCUSSION**

The results of this study show that all of the screening tests have good 1-week test-retest reliability. This means that they are all constructed in such a way as to encourage reliable reporting by adolescents over a brief interval of time. This finding is consistent with that of other studies of substance abuse self-reports, showing that measures of current use are stable over short periods, and that measures of lifetime use are stable over periods as long as 1 year.23–26 The SSI-AOD also had good

### Table 2

Distribution of Responses for Frequently Positive Items at Initial Test ($n = 173$)

<table>
<thead>
<tr>
<th>Screening Test</th>
<th>Items</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSI-AOD*</td>
<td>16 Test</td>
<td>34.0 27.8 12.5 7.6 3.5 8.3 2.8 1.4 .7 1.4 0 0 0 0 0</td>
</tr>
<tr>
<td>SSI-AOD</td>
<td>16 Retest</td>
<td>42.7 34.1 8.5 3.7 2.4 2.4 1.2 0 0 1.2 0 1.2 0 1.2 0</td>
</tr>
<tr>
<td>CAGE-AA</td>
<td>4 Test</td>
<td>70.5 17.1 7.5 4.1 .7</td>
</tr>
<tr>
<td>CAGE-AA</td>
<td>4 Retest</td>
<td>75.9 15.7 4.8 2.4 1.2</td>
</tr>
<tr>
<td>DAP-4</td>
<td>4 Test</td>
<td>43.8 29.4 18.8 6.9 1.3</td>
</tr>
<tr>
<td>DAP-4</td>
<td>4 Retest</td>
<td>51.1 24.4 21.1 2.2 1.1</td>
</tr>
</tbody>
</table>

* Suggested cut points: Total score 0–1 = Low risk; total score 2–3 = minimal risk; total score >4 = moderate to high risk.
internal consistency among all groups, indicating that individual items are likely measuring the same construct. Practitioners and researchers can therefore be assured that this screening test can be used as a scale, ie, that its total score is likely to accurately measure total level of risk. This study did not assess the sensitivity or specificity of the test and future studies should address this question. However, relatively high numbers of subjects were identified as minimal risk and moderate to high risk (20.1% and 18.1%, respectively). As a written test, the SSI-AOD may be particularly suited to busy medical offices where medical care providers find waiting room questionnaires a practical alternative to asking questions of their patients during the personal interview. When used in this fashion, however, providers should review all responses on the written test before seeing the patient and ask further questions about each positive item.

Our study found that the CAGE-AA questions have acceptable internal consistency among females, blacks and whites but unacceptable internal consistency among males and Hispanics. The precise reasons for this are unclear, and the finding of a higher α score among females is especially surprising given that Werner and colleagues12,13 found that gender had the opposite effect on validity of CAGE. However, we significantly modified the CAGE questions for use in the present study. In addition, we administered the questions in written form, whereas practitioners are usually encouraged to verbally administer them.27 It is not known how different administration strategies will affect consistency or reliability of this test, although we speculate that some adolescents may answer personal questions more honestly by way of an anonymous questionnaire than they will in a personal interview. The opposite may be true when they have a particularly good relationship with the interviewer. Further studies will be required to definitively answer this question. In any case, we believe that the CAGE test should be used cautiously with adolescent patients, and agree that it cannot stand alone as a screening instrument for AOD use. CAGE-AA should be further refined and validated before its use among adolescents can be recommended. The first step, as in this study, should be aimed at ensuring that the questions are developmentally appropriate (eg, Eye-opener question becomes “Do you ever use alcohol or drugs early in the day?”)

The developers of the DAP QuickScreen have never suggested that the 4 items with greatest discriminant validity should be used as a separate scale, and the low internal consistency found for these items has little practical meaning. We believe, however, that each is assessing a very important behavioral construct. If medical care providers could only ask 1 question of every adolescent patient, we believe it should be “Have you ever ridden in a car driven by someone, including yourself, who was high or appeared to have been using alcohol or another drug?” As stated in the introduction, alcohol-related motor vehicle crashes represent the single greatest peril to American public health, and medical care providers should actively screen every patient they see for potential risk. In fact, we have just published a study showing that a new brief test, which contains this question, is strongly correlated with the need for referral to substance abuse treatment.28 Clinicians must be prepared to offer immediate brief advice and counseling when patients answer this question affirmatively.

This study also underscores the importance of routine screening for AOD, however it is accomplished. The 6-month prevalence of AOD-related risk and problem behaviors is high among our adolescent clinic patients. Six percent of our subjects had experienced a very serious problem as a result of AOD use, yet few of them had asked for help. Clinicians should not therefore expect that their patients will come forward and request counseling or referral on their own. They must routinely ask about AOD-related problems so that early intervention and referral can be offered.

This study has limitations. It was conducted among a group of adolescent medical patients, and the generalizability of the results to the adolescent population at-large is unknown. The study measured reliability but not validity of the screening instruments and, in fact, the CAGE and DAP questions were significantly modified from their previously validated forms. We must therefore refrain from making judgments about the psychometric properties of either original instrument until further studies are completed.

### CONCLUSION

The SSI-AOD has good internal consistency and test-retest reliability. Further studies must confirm its psychometric properties, but it appears promising as a reliable screening measure of substance abuse risk among adolescents seen in medical settings. The CAGE-AA questions should be further modified and tested before their use can be recommended in adolescents. The DAP-4 questions are likely measuring different, but important, constructs. Regardless of the specific test that is used,
however, routine screening of all adolescents for AOD use must be the standard of care in medical office practice.

ACKNOWLEDGMENTS

This work was supported in part by Grant No. MCJMA259195 and Grant No. MCJ2550360 from the Maternal and Child Health Bureau Health Resources and Services Administration, US Department of Health and Human Services.

We wish to thank S. Jean Emans, MD, and Judith S. Palfrey, MD, (Children’s Hospital) for reviewing the manuscript; and Lon R. Sherritt, MPH, for advice on statistical tests.

REFERENCES

Reliabilities of Short Substance Abuse Screening Tests Among Adolescent Medical Patients
John R. Knight, Elizabeth Goodman, Todd Pulerwitz and Robert H. DuRant

Updated Information & Services
Updated Information & Services including high resolution figures, can be found at:
/content/105/Supplement_3/948.full.html

References
This article cites 18 articles, 1 of which can be accessed free at:
/content/105/Supplement_3/948.full.html#ref-list-1

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
Substance Abuse
/cgi/collection/substance_abuse_sub

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
/site/misc/Permissions.xhtml

Reprints
Information about ordering reprints can be found online:
/site/misc/reprints.xhtml
Reliabilities of Short Substance Abuse Screening Tests Among Adolescent Medical Patients
John R. Knight, Elizabeth Goodman, Todd Pulerwitz and Robert H. DuRant

Pediatrics 2000;105;948

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/105/Supplement_3/948.full.html