Reliability and Validity of the Newton Screen for Alcohol and Cannabis Misuse in a Pediatric Emergency Department Sample

James G. Linakis, PhD, MD[1,2,3], Julie R. Bromberg, MPH[1,3], T. Charles Casper, PhD[4], Thomas H. Chun, MD, MPH[1,2,3], Michael J. Mello, MD, MPH[1,3], Hailey Ingebretsen, MStat[4], and Anthony Spirito, PhD[5], for the Pediatric Emergency Care Applied Research Network

Objectives To determine the test-retest reliability, concurrent, convergent, and discriminant validity of a recently devised screen (the Newton screen) for alcohol and cannabis use/misuse, and its predictive validity at follow-up.

Study design Adolescents, 12-17 years old (n = 4898), treated in 1 of 16 participating pediatric emergency departments across the US were enrolled in a study as part of a larger study within the Pediatric Emergency Care Applied Research Network. Concurrent and predictive validity (at 1, 2, and 3 years of follow-up) were assessed in a random subsample with a structured *Diagnostic and Statistical Manual of Mental Disorders*-based interview. Convergent validity was assessed with the Alcohol Use Disorders Identification, a widely used alcohol screening measure.

Results The sensitivity of the Newton screen for alcohol use disorder at baseline was 78.3% with a specificity of 93.0%. The cannabis use question had a baseline sensitivity of 93.1% and specificity of 93.5% for cannabis use disorder. Predictive validity analyses at 1, 2, and 3 years revealed high specificity but low sensitivity for alcohol and high specificity and moderate sensitivity for cannabis.

Conclusions The Newton screening instrument may be an appropriate brief screening tool for use in the busy clinical environment. Specificity was high for both alcohol and cannabis, but sensitivity was higher for cannabis than alcohol. Like other brief screens, more detailed follow-up questions may be necessary to definitively assess substance misuse risk and the need for referral to treatment. *(J Pediatr 2019;210:154-60)*.

In 2016, 7.3 million 12- to 20-year-olds reported using alcohol in the past 30 days, with 62.5% of these individuals reporting binge alcohol use (≥5 drinks per occasion for males; ≥4 drinks per occasion for females) and 14.7% reporting heavy alcohol use (binge alcohol use ≥5 times in the past 30 days).[1] In 2013, 1.3 million adolescents (5.2% of adolescents) aged 12-17 years met the criteria for a substance use disorder.[2] Adolescents who use alcohol and drugs have been found to be more likely to participate in other risky behaviors and experience negative health consequences compared with their peers who do not use substances.[3,4] Further, early initiation of alcohol use is also associated with alcohol misuse and more severe alcohol problems in the future.[5,6] This level of adolescent substance use contributes to high levels of alcohol- and drug-related pediatric emergency department (PED) visits. The Drug Abuse Warning Network estimates that in 2011 there were about 280 000 drug-related ED visits by adolescents aged 12-17 years.[7] Alcohol and drug screening is necessary to appropriately identify patients and intervene (when appropriate) with a brief intervention and/or referral to treatment. Universal alcohol and drug screening in PEDs is recommended by several federal organizations.[8-10]

Numerous options for alcohol and/or drug screening exist for adolescents.[11] However, current instruments are often challenging for busy clinicians to administer. To be suitable for use in a variety of clinical settings, screening instruments must require minimal training and implementation time. Recently, several screening tools were introduced for use in adolescents. The National Institute for Alcohol Abuse and Alcoholism (NIAAA) disseminated a 2-question alcohol

AUD	Alcohol use disorder
AUDIT	Alcohol Use Disorder Identification Test
CUD	Cannabis use disorder
DISC	Diagnostic Interview Schedule for Children
DSM	*Diagnostic and Statistical Manual of Mental Disorders*
NIAAA	National Institute for Alcohol Abuse and Alcoholism
PECARN	Pediatric Emergency Care Applied Research Network
PED	Pediatric emergency department
YRBBS	Youth Risk Behavior Surveillance System

[1]The Warren Alpert Medical School of Brown University, Emergency Medicine, Providence, RI; [2]The Warren Alpert Medical School of Brown University, Pediatrics, Providence, RI; [3]Rhode Island Hospital, Emergency Medicine, Providence, RI; [4]University of Utah, Pediatrics, Salt Lake City, UT; and [5]The Warren Alpert Medical School of Brown University, Psychiatry and Human Behavior, Providence, RI

All phases of this study were supported in part by NIAAA (1R01AA021900 [to A.S. and J.L.]). This project is supported in part by the Health Resources and Services Administration (HRSA), Maternal and Child Health Bureau (MCHB), Emergency Medical Services for Children (EMSC) Network Development Demonstration Program under cooperative agreements U03MC00008 and U03MC00001, U03MC00003, U03MC00006, U03MC00007, U03MC22684, and U03MC22685. This information or content and conclusions are those of the author and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS or the U.S. Government. The authors declare no conflicts of interest.

https://doi.org/10.1016/j.jpeds.2019.02.038

screen containing 1 question about the patient's frequency of drinking and a second about the patient's friends' alcohol use.[12] The initial reliability and validity for the NIAAA 2-question alcohol screen has been documented for use in the PED.[13] A positive screen suggests the need for further for alcohol use disorder (AUD).

The Screening to Brief Intervention tool uses a single screening question to assess past-year frequency of use for 8 categories of frequently misused substances and has been shown to be a valid instrument for assessing risk categories of adolescent substance use.[14]

A review by Newton et al recommended using 3 questions to screen for adolescent substance use based on the *Diagnostic and Statistical Manual of Mental Disorders-Version 5* (DSM-5) of the substance use disorder module of the Diagnostic Interview Schedule for Children (DISC).[15,16] The questions are: "1. In the past year, have you sometimes been under the influence of alcohol in situations where you could have caused an accident or gotten hurt? 2. Have there often been times when you had a lot more to drink than you intended to have? and 3. In the past year, how often have you used cannabis: 0 to 1 time, ≥2 times?" This screening instrument, the "Newton screen," was specifically proposed for use in the PED. However, it has not, to our knowledge, been validated in any setting. Understanding the performance characteristics of this screen is important for providers to understand as they select and use substance use screening tools. The objective of this study was to determine the test-retest reliability and concurrent, convergent, and discriminant validity of the Newton screen and its predictive validity at the 1-, 2-, and 3-year follow-up visits.

Methods

This was a prospective observational study of a convenience sample of adolescents seen in 1 of 16 PEDs. After obtaining parental permission and patient assent, a criterion assessment battery was self-administered on a tablet computer by participants 12-17 years of age within 16 PEDs belonging to the Pediatric Emergency Care Applied Research Network (PECARN) across a 25-month period from May 2013 to June 2015. The battery was self-administered on a tablet computer and included the NIAAA 2-question screen, the Newton screen, and other measures of alcohol, drug use, and risk behavior, including number of drinking days in past year, and the DISC, a structured, DSM-based interview used to determine alcohol use diagnoses, and the Alcohol Use Disorder Identification Test (AUDIT).[12,15,17] All sites received institutional review board approval before conducting research activities. Owing to the potential legal implications of illicit alcohol or drug use, a Certificate of Confidentiality was obtained from the US Department of Health and Human Services. The research protocol has been described elsewhere.[18]

To determine test-retest reliability, a random sample of enrolled participants was contacted by phone and email 7-14 days after the PED visit to complete the Newton screen for a second time. A subsample of our original sample was randomly selected to be readministered the Newton screen and the criterion assessment battery at 12, 24, and 36 months after their ED visit to examine the predictive validity of the Newton screen.

Key Outcome Measures

The Newton screen consists of 2 DSM-5 screen questions for detecting youth alcohol misuse (alcohol influence in dangerous situations and frequently drinking more than intended). The Newton screen is considered positive for alcohol misuse when the participant responds in the affirmative to either alcohol question.[15] The screen also includes a question about cannabis use in the past year, with response categories of 0-1 time and 2 or more times. A response of 2 or more times is considered a positive cannabis screen.[15]

Concurrent validity, the degree to which the results of a test are comparable with those of an established, gold standard measure of the same construct, was assessed with the Alcohol- and Substance-Use Disorder modules of the DISC, a structured DSM-based interview.[17] DSM-5 diagnoses for AUD and cannabis use disorder (CUD) were derived based on participant responses on the DISC. Convergent validity for alcohol use, the degree to which 2 tests designed to assess the same construct are related, was assessed with the AUDIT, a 10-question screen focusing on the quantity and frequency of alcohol use and frequently used to detect AUDs in both adults and adolescents.[19] Discriminant validity (the degree to which constructs that are not supposed to be related are actually unrelated) was assessed using one item of the Youth Risk Behavior Surveillance System (YRBSS) physical activity subscale.[20] Predictive validity was assessed by readministering the Newton questions and the criterion battery 1, 2, and 3 years after the index PED visit.

Statistical Analyses

The sensitivity for all of the screening instruments was used as the basis for sample size requirements. We assumed a target sensitivity of 90%. For the 95% CI around sensitivity to be within ±2.5%, we targeted an enrollment of approximately 5000 participants. We determined that approximately 200 participants with a 1-week follow-up would provide a stable estimate of test-retest reliability.

For analyses regarding the validity of the alcohol misuse portion of the Newton screen, participants were included if they completed the 2 Newton alcohol misuse questions. Participants completing the cannabis use question of the Newton screen were included in analyses related to cannabis use. For test-retest reliability analyses, we included participants who were assigned to a 1-week follow-up and completed the Newton screen items at both baseline and the 1-week follow-up. The Cohen kappa statistic was used to estimate the reliability between positive Newton screen for alcohol and cannabis at baseline and the 1-week follow-up.

Concurrent and predictive validity were analyzed using sensitivity, specificity, and predictive values of the alcohol portion of the Newton screen when predicting the DISC diagnostic outcome of AUD (based on DSM-5 criteria). We performed a similar analysis for the DISC diagnostic outcome of CUD using the Newton screen cannabis item. We also conducted logistic regression using the Newton Screen to predict a positive AUD and CUD diagnosis, respectively, at baseline and at follow-up.

To examine convergent validity, the distributions of AUDIT total scores between positive and negative Newton screens for alcohol were compared using the Wilcoxon rank-sum test. We defined an AUDIT cutoff of 4 or greater as high risk, because this cutoff has been used in research on adolescents.[21] For this AUDIT high-risk outcome, we analyzed the performance of the Newton screen with respect to sensitivity, specificity, and predictive values. We did not conduct convergent validity analysis with respect to cannabis use because there is currently no accepted screen for cannabis. To assess the discriminant validity of the Newton screen—examining if the Newton screen results were unrelated to YRBSS physical activity results (because the 2 screens are not designed to assess the same construct)—the Wilcoxon rank-sum test was used to compare the single item of the YRBSS physical activity subscale across the Newton screen outcomes for alcohol and cannabis misuse separately. All analyses were performed using SAS version 9.4 (SAS Institute, Cary, North Carolina).

Results

A total of 4898 participants completed both alcohol items or the cannabis item of the Newton screen at baseline in the PED (Figure). A total of 190 of the 276 participants (69%) assigned to the follow-up group completed both the baseline and 1-week follow-up assessments of the Newton screen. The 1-, 2-, and 3-year follow-up rates were 71%, 69%, and 60%, respectively. The sociodemographic characteristics of the full sample are reported in Bromberg et al.[19] Table I provides the demographic breakdown by Newton screen status at baseline.

Test-Retest Reliability

The Newton screen was compared at baseline vs the 1-week follow-up for the 190 participants who provided data at both time points. Participants were asked if they drank alcohol in the past week to account for any true alcohol consumption as opposed to retest error. Three participants responded "yes" to a question about whether they had a drink in the past week at the 1-week follow-up. The analyses were essentially unchanged when these participants were dropped from the analyses. Consequently, the results for the entire sample are reported here. The alcohol screen produced a kappa coefficient of 0.64 (95% CI, 0.43-0.84) and the cannabis question a kappa of 0.75 (95% CI, 0.58-0.93), representing moderate agreement.[22]

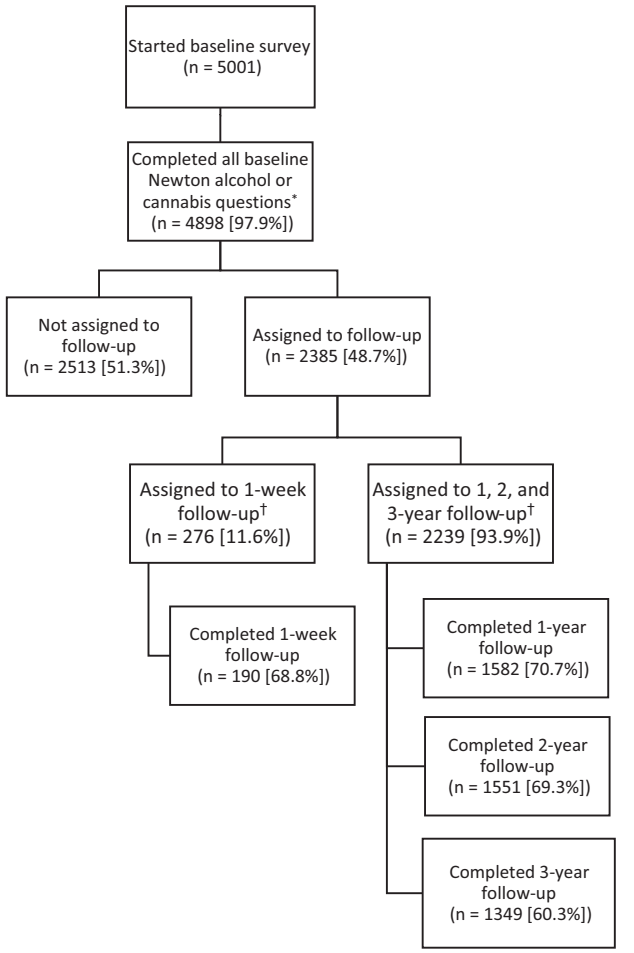

Figure. Enrollment diagram. *There were 4677 participants who completed all baseline Newton alcohol questions; 4525 completed the baseline Newton cannabis question and 4304 completed all baseline Newton alcohol and cannabis questions. †Includes those assigned to both 1-week and 1-, 2-, and 3-year follow-up groups.

Concurrent and Predictive Validity

Table II presents the breakdown of AUD and CUD diagnosis on the DISC by positive Newton alcohol screen at baseline (concurrent validity) and at 1, 2, and 3 years of follow-up (predictive validity). The sensitivity, specificity, and positive and negative predictive values of the Newton screen for AUD and CUD diagnosis at baseline and 1, 2, and 3 years of follow-up are shown in Table III. At baseline, the Newton alcohol screen had high specificity 93.0% (95% CI, 92.3%-93.8%), but only moderate sensitivity 78.3% (95% CI, 68.5%-88.0%). At follow-up, specificity remained high, but sensitivity became poor. Inspection of the alcohol questions individually at baseline revealed that the question regarding drinking more than intended had higher sensitivity, 72.6% (95% CI, 62.4%-82.8%) than the dangerous situations question, 46.4% [95% CI, 34.6%-58.1%]). Both sensitivity, 93.1% (95% CI, 90.1%-96.1%) and specificity, 93.5% (95% CI 92.7%-

Table I. Participant demographics and baseline characteristics

Demographic group	Overall (n = 4898)	Baseline Newton screen for alcohol misuse		Baseline Newton screen for cannabis misuse	
		Positive (n = 385)	Negative (n = 4292)	Positive (n = 553)	Negative (n = 3972)
Age, y	14.5 ± 1.66	15.4 ± 1.44	14.4 ± 1.65	15.6 ± 1.31	14.4 ± 1.64
Age group					
Middle school	1901 (38.8)	59 (15.3)	1757 (40.9)	70 (12.7)	1624 (40.9)
High school	2997 (61.2)	326 (84.7)	2535 (59.1)	483 (87.3)	2348 (59.1)
Sex					
Male	2215 (45.2)	160 (41.6)	1966 (45.8)	247 (44.7)	1788 (45.0)
Female	2683 (54.8)	225 (58.4)	2326 (54.2)	306 (55.3)	2184 (55.0)
Ethnicity					
Hispanic or Latino	1278 (26.1)	111 (28.8)	1115 (26.0)	139 (25.1)	1011 (25.5)
Not Hispanic or Latino	3408 (69.6)	260 (67.5)	2992 (69.7)	397 (71.8)	2793 (70.3)
Unknown or not reported	212 (4.3)	14 (3.6)	185 (4.3)	17 (3.1)	168 (4.2)
Race					
White	2264 (46.2)	188 (48.8)	1995 (46.5)	209 (37.8)	1959 (49.3)
Black	1293 (26.4)	84 (21.8)	1134 (26.4)	186 (33.6)	973 (24.5)
American Indian/Alaska Native	98 (2.0)	11 (2.9)	81 (1.9)	13 (2.4)	75 (1.9)
Asian	62 (1.3)	4 (1.0)	51 (1.2)	3 (0.5)	56 (1.4)
Native Hawaiian or other Pacific Islander	44 (0.9)	5 (1.3)	37 (0.9)	5 (0.9)	35 (0.9)
More than one race	375 (7.7)	32 (8.3)	327 (7.6)	51 (9.2)	299 (7.5)
Unknown or not reported	762 (15.6)	61 (15.8)	667 (15.5)	86 (15.6)	575 (14.5)

Numerical variables are summarized as mean ± SD and categorical variables summarized as number (%).

94.2%) of the Newton screen for CUD were high at baseline but decreased each year. The positive predictive values of the Newton screen for AUD was similar at baseline (14.6%) and the 3 follow-up time points (10.7% at 1 year, 12.4% at 2 years, and 16.7% at 3 years). The positive predictive values of the screen for CUD were also similar at baseline (48.7%) and follow-up (38.6% at 1 year, 41.2% at 2 years, and 43.6% at 3 years) yet higher than those for AUD across all time periods.

Convergent Validity of the Alcohol Screen

Table IV presents the AUDIT scores by positive Newton alcohol screens. The overall test comparing Newton screen positive vs negative was statistically significant ($P < .001$). Of the participants who were positive on the Newton screen, 50.9% had an AUDIT score of 4 or higher. The sensitivity of the screen for predicting an AUDIT score of 4 or higher was 59.7% (95% CI, 54.2%-65.2%), with a specificity of 95.8% (95% CI, 95.2%-96.4%). Positive and negative predictive values were 51.0% (95% CI, 45.8%-56.2%) and 97.0% (95% CI, 96.5%-97.6%), respectively. Again, the second question (drinking more than intended) was more predictive of a high AUDIT score than the first question, with a sensitivity and specificity of 53.9% (95% CI, 48.4%-59.3%) and 97.4% (95% CI, 97.0%-97.9%), respectively, compared with 25.3% (95% CI, 20.5%-30.2%) and 97.7% (95% CI, 97.2%-98.1%) for the first question.

Discriminant Validity

A Wilcoxon rank-sum test comparing 1 item of the YRBSS physical activity subscale across each of the Newton screen classifications and questions was not significant for positive screens on the alcohol questions ($P = .45$) or for the cannabis question ($P = .81$).

Table II. Distribution of DSM-5 diagnosis of AUD or CUD at baseline and follow-up by baseline Newton Screen results

DSM-5 diagnoses	Baseline Newton screen for alcohol misuse			
	Positive	Negative	*P* value*	OR (95% CI)
AUD				
Baseline	54/369 (14.6%)	15/4230 (0.4%)	<.0001	48.17 (26.88-86.32)
1 year	12/112 (10.7%)	17/1392 (1.2%)	<.0001	9.71 (4.51-20.89)
2 years	13/105 (12.4%)	32/1362 (2.3%)	<.0001	5.87 (2.98-11.57)
3 years	13/78 (16.7%)	43/1193 (3.6%)	<.0001	5.35 (2.74-10.44)
CUD				
Baseline	257/528 (48.7%)	19/3901 (0.5%)	<.0001	193.8 (119.7-313.7)
1 year	53/144 (36.8%)	34/1310 (2.6%)	<.0001	21.86 (13.52-35.34)
2 years	56/136 (41.2%)	51/1292 (3.9%)	<.0001	17.03 (10.95-26.49)
3 years	44/101 (43.6%)	85/1137 (7.5%)	<.0001	9.56 (6.09-15.00)

Fractions represent the number of participants classified as having DSM-5 AUD or CUD at each time point over the number classified as positive or negative by the Newton screen. For example, at baseline, of the 369 classified as positive for alcohol misuse by the Newton screen, 54 were classified as positive for AUD by the DSM-5.
**P* values are based on a logistic regression model with Newton Screen predicting a positive AUD or CUD diagnosis at baseline and follow-up.

Table III. Sensitivity and specificity of a positive Newton screen for alcohol misuse or cannabis misuse predicting DSM-5 diagnosis of AUD or CUD at baseline and follow-up

DSM-5 diagnoses	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
AUD				
Baseline	0.783 (0.685-0.880)	0.930 (0.923-0.938)	0.146 (0.110-0.182)	0.996 (0.995-0.998)
1-year follow-up	0.414 (0.235-0.593)	0.932 (0.919-0.945)	0.107 (0.050-0.164)	0.988 (0.982-0.994)
2-year follow-up	0.289 (0.156-0.421)	0.935 (0.923-0.948)	0.124 (0.061-0.187)	0.977 (0.968-0.985)
3-year follow-up	0.232 (0.122-0.343)	0.947 (0.934-0.959)	0.167 (0.084-0.249)	0.964 (0.953-0.975)
CUD				
Baseline	0.931 (0.901-0.961)	0.935 (0.927-0.942)	0.487 (0.444-0.529)	0.995 (0.993-0.997)
1-year follow-up	0.609 (0.507-0.712)	0.933 (0.920-0.947)	0.368 (0.289-0.447)	0.974 (0.965-0.983)
2-year follow-up	0.523 (0.429-0.618)	0.939 (0.927-0.952)	0.412 (0.329-0.494)	0.961 (0.950-0.971)
3-year follow-up	0.341 (0.259-0.423)	0.949 (0.936-0.962)	0.436 (0.339-0.532)	0.925 (0.910-0.941)

NPV, Negative predictive value; *PPV*, Positive predictive value.

Discussion

The American Academy of Pediatrics, the American College of Surgeons, and other organizations endorse universal screening of youth presenting to the ED for alcohol and substance misuse, and such screening is frequently recommended at primary care visits.[8-10,12] Partly as a result of time pressures under which emergency and primary care physicians operate, existing screening instruments are often impractical.[23,24] A recent report recommends 3 potential options for adolescent brief alcohol screens: the Screening to Brief Intervention, Brief Screener for Tobacco, Alcohol, and other Drugs, and the NIAAA 2-question screen.[12,14,25,26] We recently completed a study of the NIAAA 2-question screen and found it to be both a valid and brief alternative to other screening instruments.[13] The present study examined the reliability and validity of another potential brief screening option, the Newton et al screening instrument, in the emergency department setting using a multicenter design. The study sample was substantively larger than previous studies designed to test validity of alcohol screening instruments in adolescents, and was racially, ethnically, and geographically diverse.[13,18,21,27-29]

The test-retest reliability of the screen was adequate for both the alcohol misuse component and the cannabis question. The reliability of the overall alcohol misuse screen (≥1 items positive) was acceptable. However, when the 2 alcohol questions were analyzed separately, responses to the dangerous situation alcohol misuse question were less reliable than responses to the drinking more than intended question. Although the explanation for this difference is unclear, it is possible that the dangerous situation question is conceptually more difficult for teens to understand, thereby compromising the accuracy of their responses. This factor could also explain the finding that this question had notably less sensitivity when assessing validity than the drinking more than intended question. This study suggests that the drinking more than intended question is the stronger of the 2 alcohol questions and that it may be appropriate to use independently as a quick alcohol screen.

Table IV. Convergent validity: Distribution of AUDIT scores by Newton screen for alcohol misuse at baseline

AUDIT total scores at baseline	Baseline Newton screen for alcohol misuse	
	Positive (n = 357)	Negative (n = 4157)
0	94 (26.3%)	3635 (87.4%)
1-3	81 (22.7%)	399 (9.6%)
4	28 (7.8%)	45 (1.1%)
>4	154 (43.1%)	78 (1.9%)

AUDIT scores are presented as number (%).
Excluding 163 participants with missing baseline AUDIT scores.

When the instrument was validated against the DISC, the Newton screen alcohol questions proved less sensitive than the marijuana question, but still demonstrated good sensitivity and excellent specificity. Indeed, when used as suggested, 78% of those in our sample who were diagnosed with an AUD on the DISC were successfully identified by the Newton screen. It is important to note, however, that the screening instrument had a modest positive predictive value of 15%, suggesting that, when used as intended, there are likely to be a significant number of individuals who have a false-positive screen. Because the proposed value of this screening instrument is to permit quick screening of youth in the emergency department or other clinical setting, a positive result on the screen would likely warrant further evaluation to verify the risk level of the adolescent who screens positive. With respect to convergent validity, when compared with a frequently used alcohol screen (AUDIT), the sensitivity of the Newton screen in comparison with an AUDIT score of 4 or greater—a value frequently used as a clinical cut-off for adolescents—the sensitivity was 59.7% with a positive predictive value of 51.0%.[21,28] These data suggest a fair but not outstanding performance for the Newton screen compared with the AUDIT. Importantly, there are other test characteristics to consider in addition to performance when selecting an appropriate screening tool for the PED. The Newton screen addresses alcohol and other drugs, consists of 3 questions, and is easy to administer and score, whereas the AUDIT consists of 10 alcohol-related questions and is more complicated to score.

As with all screens, if a teen screens positive then more in-depth follow-up questions, such as questions about craving,

withdrawal, use patterns, and effects on school performance, are necessary to determine whether further evaluation/referral to treatment is indicated in an individual case. The specific protocol for using the instrument optimally in the PED remains to be developed and tested. For example, although our concurrent validity analysis was intended to identify those with an actual AUD, a more desirable goal in the clinical setting might be to identify those whose drinking is problematic but has not yet attained the level of an AUD. Thus, a positive response to the Newton screen could be used to trigger a more in-depth inquiry into substance use by interview or a more comprehensive self-report measure. One approach, for example, would be to use the 2 alcohol questions from the Newton screen to triage patients to low risk or further testing and then to use a more comprehensive assessment as the next step to aid decisions about further evaluation and referral to treatment. The nature of the more comprehensive assessment remains to be defined, but a promising approach is computerized adaptive testing, which has the advantage of asking a teen only those assessment items that are relevant to their specific level of substance use.[30]

A comparison of the Newton alcohol screen and the NIAAA 2-question screen with respect to test-retest reliability and concurrent validity reveals the 2 instruments to be quite similar.[13] The NIAAA screen had a kappa coefficient of 0.63, whereas the alcohol component of the Newton screen had a kappa coefficient of 0.64.[13] Although concurrent validity was measured differently for the 2 screens because of differences in the way that the respective instruments are intended to be administered, the 2 screening instruments were comparable. The Newton screen had a sensitivity of 78% (across all ages) and a specificity of 93%, whereas the NIAAA screen had a sensitivity of 89% for middle school and 88% for high school, and a specificity of 91% and 81% for middle school and high school, respectively.[13]

For cannabis misuse, the Newton screen has a single question that has reasonably high positive predictive value (0.49) for a DISC-diagnosed CUD at baseline. Given its high negative predictive value, this screen might prove to be feasible in the clinical setting and, when positive, could be used by the clinician to explore further with the patient the need for referral to treatment. As mentioned, the screen's test characteristics seem to be better for cannabis than for alcohol.

Finally, the predictive validity portion of our results revealed a progressively decreasing sensitivity for both the alcohol and cannabis portion of the Newton screen with time, although this was more striking for the alcohol questions. Even given this decreasing sensitivity, the benefits of this short screen may support its implementation. The specificity remained robust for both portions, as did the negative predictive value. The positive predictive value remained moderate over time for the cannabis portion of the screen, whereas for the alcohol portion it remained low. As with the concurrent usefulness of the screen, the predictive usefulness of the Newton screen remains to be established. Certainly, in most clinical settings, the vast majority of youth will screen negative, and in light of the screen's negative predictive value, it is unlikely that further screening would be warranted. However, for those screening positive, it could be anticipated that the majority, especially when predicting future risk, will represent false positives. The means for discerning those who are, in fact, at risk for developing substance use disorders over time remains to be clarified.

There are a number of limitations to our study. First, the test items on the Newton screen were derived from (but are not identical to) the DISC, and the DISC was used to validate the Newton screen. This factor may have affected our results owing to the relationship between the Newton screen and the DISC, although the effect is likely to be small because we were comparing the screening instrument to specific diagnoses from the DISC rather than a total score of the DISC items. In addition, our findings on sensitivity and specificity were not of such a magnitude to suggest this methodologic issue seemed to skew the findings. Second, although the sample was both large and demographically diverse, it was limited to adolescents being evaluated in PECARN PEDs and, therefore, was not representative of the general population. Third, the administration of one screen measuring the same factor as the other screen may have influenced the results of each one. Similarly, although the order of the criterion measures was varied when the survey was administered, the Newton screen was always administered after the NIAAA screen. Both could have conceivably had an effect on participant response style. Fourth, surveys were all self-administered, and thus we cannot comment on how this screening instrument would perform when the questions were asked directly by a healthcare provider.

Although the screening instrument proposed by Newton et al represents a feasible approach to brief screening in the busy clinical environment, further refinement of screening strategies is required. Additional research is required to determine the optimal combinations of screening instruments and screening strategies in the PED setting, while preserving optimal efficiency and minimal disruption of normal PED care. ■

*We thank the PECARN Steering Committee members, staff at the Data Coordinating Center, and the other participating PECARN site investigators (**Appendix**).*

Submitted for publication Oct 15, 2018; last revision received Feb 15, 2019; accepted Feb 26, 2019.

References

1. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2016 National Survey on Drug Use and Health. Rockville (MD): Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2017.
2. Substance Abuse and Mental Health Services Administration. Substance Use and Mental Health Estimates from the 2013 National Survey on

Drug Use and Health: overview of Findings. Rockville, (MD): Substance Abuse and Mental Health Services Administration; 2014.
3. Windle M, Zucker RA. Reducing underage and young adult drinking: how to address critical drinking problems during this developmental period. Alcohol Res Health 2010;33:29-44.
4. Bohnert KM, Walton MA, Resko S, et al. Latent class analysis of substance use among adolescents presenting to urban primary care clinics. AM J Drug Alcohol Abuse 2014;40:44-50.
5. Grant BF, Stinson FS, Harford TC. Age at onset of alcohol use and DSM-IV alcohol abuse and dependence: a 12-year follow-up. J Subst Abuse 2001;13:493-504.
6. Hawkins JD, Graham JW, et al. Exploring the effects of age of alcohol use initiation and psychosocial risk factors on subsequent alcohol misuse. J Stud Alcohol 1997;58:280-90.
7. Substance Abuse and Mental Health Services Administration. The CBHSQ Report: a day in the life of American adolescents: substance use facts update. Rockville, (MD): Substance Abuse and Mental Health Services Administration; 2013.
8. American College of Emergency Physicians. Alcohol screening in the emergency department policy statement., http://www.acep.org/Content.aspx?id=29074. [Accessed 30 June 2016].
9. Committee on Substance Use and Prevention. Substance use screening, brief intervention, and referral to treatment. Pediatrics 2016;138:e20161210.
10. Substance Abuse and Mental Health Services Administration. SBIRT: Screening, Brief Intervention, and Referral to Treatment: opportunities for implementation and points for consideration., http://www.integration.samhsa.gov/SBIRT_Issue_Brief.pdf. [Accessed 10 June 2016].
11. Knight JR, Shrier LA, Bravender TD, Farrell M, Vander Bilt J, Shaffer HJ. A new brief screen for adolescent substance abuse. Arch Pediatr Adolesc Med 1999;153:591-6.
12. National Institute on Alcohol Abuse and Alcoholism. Alcohol screening and brief intervention for youth: a practitioner's guide. 2011, www.niaaa.nih.gov/YouthGuide. [Accessed 20 April 2016].
13. Spirito A, Bromberg JR, Casper TC, et al. Reliability and validity of a two-question alcohol screen in the pediatric emergency department. Pediatrics 2016;138.
14. Levy S, Weiss R, Sherritt L, et al. An electronic screen for triaging adolescent substance use by risk levels. JAMA Pediatr 2014;168:822-8.
15. Newton AS, Gokiert R, Mabood N, et al. Instruments to detect alcohol and other drug misuse in the emergency department: a systematic review. Pediatrics 2011;128:e180-92.
16. Shaffer D, Fisher P, Dulcan MK, et al. The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3): description, acceptability, prevalence rates, and performance in the MECA Study. Methods for the Epidemiology of Child and Adolescent Mental Disorders Study. J Am Acad Child Adolesc Psychiatry 1996;35:865-77.
17. Fisher PW, Shaffer D, Piacentini JC, et al. Sensitivity of the Diagnostic Interview Schedule for Children, 2nd edition (DISC-2.1) for specific diagnoses of children and adolescents. J Am Acad Child Adolesc Psychiatry 1993;32:666-73.
18. Bromberg J, Spirito A, Chun T, et al. Methodology and demographics of a brief adolescent alcohol screen validation study. Pediatr Emerg Care 2017, in press.
19. Fairlie AM, Sindelar HA, Eaton CA, Spirito A. Utility of the AUDIT for screening adolescents for problematic alcohol use in the emergency department. Int J Adolesc Med Health 2006;18:115-22.
20. Centers for Disease Control and Prevention. 2013 State and Local Youth Risk Behavior Survey. 2013, http://www.cdc.gov/healthyyouth/yrbs/pdf/questionnaire/2013_hs_questionnaire.pdf. [Accessed 13 April 2015].
21. Chung T, Colby SM, Barnett NP, Rohsenow DJ, Spirito A, Monti PM. Screening adolescents for problem drinking: performance of brief screens against DSM-IV alcohol diagnoses. J Stud Alcohol 2000;61:579-87.
22. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37-46.
23. McNeely J, Kumar PC, Rieckmann T, et al. Barriers and facilitators affecting the implementation of substance use screening in primary care clinics: a qualitative study of patients, providers, and staff. Addict Sci Clin Pract 2018;13:8.
24. Sterling S, Kline-Simon AH, Wibbelsman C, Wong A, Weisner C. Screening for adolescent alcohol and drug use in pediatric health-care settings: predictors and implications for practice and policy. Addict Sci Clin Pract 2012;7:13.
25. Levy SJ, Williams JF , Committee on Substance Use and Prevention. Substance use screening, brief intervention, and referral to treatment. Pediatrics 2016;138.
26. Kelly SM, Gryczynski J, Mitchell SG, Kirk A, O'Grady KE, Schwartz RP. Validity of brief screening instrument for adolescent tobacco, alcohol, and drug use. Pediatrics 2014;133:819-26.
27. Chung T, Colby SM, Barnett NP, Monti PM. Alcohol use disorders identification test: factor structure in an adolescent emergency department sample. Alcohol Clin Exp Res 2002;26:223-31.
28. Knight JR, Sherritt L, Harris SK, Gates EC, Chang G. Validity of brief alcohol screening tests among adolescents: a comparison of the AUDIT, POSIT, CAGE, and CRAFFT. Alcohol Clin Exp Res 2003;27:67-73.
29. Knight JR, Sherritt L, Shrier LA, Harris SK, Chang G. Validity of the CRAFFT substance abuse screening test among adolescent clinic patients. Arch Pediatr Adolesc Med 2002;156:607-14.
30. Gibbons RD, Weiss DJ, Frank E, Kupfer D. Computerized adaptive diagnosis and testing of mental health disorders. Annu Rev Clin Psychol 2016;12:83-104.

Appendix

Additional members of the Pediatric Emergency Care Applied Research Network (PECARN)

Rohit P. Shenoi, MD, Baylor College of Medicine/Texas Children's Hospital, Houston, TX

Fahd Ahmad, MD, MSCI, St. Louis Children's Hospital/Washington University, St. Louis, MO

Lalit Bajaj, MD, MPH, Children's Hospital–Colorado, Aurora, CO

Kathleen M. Brown, MD, Children's National Medical Center, Washington, DC

Lauren S. Chernick, MD, MSc, Columbia University Medical Center, New York, NY

Daniel M. Cohen, MD, Nationwide Children's Hospital, Columbus, OH

J. Michael Dean, MD, MBA, University of Utah, Salt Lake City, UT

Joel Fein, MD, MPH, The Children's Hospital of Philadelphia, Philadelphia, PA

Jackie Grupp-Phelan, MD, MPH, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

Timothy Horeczko, MD, MSCR, FACEP, FAAP, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Los Angeles, CA

Michael N. Levas, MD, Medical College of Wisconsin, Milwaukee, WI

B. McAninch, MD, University of Pittsburgh/Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA

Michael C. Monuteaux, ScD, Boston Children's Hospital, Boston, MA

Colette C. Mull, MD, Sidney Kimmel Medical College at Jefferson University/Nemours Alfred I. duPont Hospital for Children, Wilmington, DE

Elizabeth C. Powell, MD, MPH, Lurie Children's Hospital of Chicago, Chicago, IL

Alexander Rogers, MD, University of Michigan, Ann Arbor, MI

Brian Suffoletto, MD, MS, University of Pittsburgh/Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA

Cheryl Vance, MD, University of California, Davis, Davis, CA.