


RESEARCH

Open Access



A systematic review of validity of US survey measures for assessing substance use and substance use disorders

Yuni Tang^{1,4}, Erin Caswell¹, Rowida Mohamed^{2,5}, Natalie Wilson³, Edis Osmanovic³, Gordon Smith¹, Summer D. Hartley⁶ and Ruchi Bhandari^{1*} 

Abstract

Background The steep rise in substance use and substance use disorder (SUD) shows an urgency to assess its prevalence using valid measures. This systematic review summarizes the validity of measures to assess the prevalence of substance use and SUD in the US estimated in population and sub-population-based surveys.

Methods A literature search was performed using nine online databases. Studies were included in the review if they were published in English and tested the validity of substance use and SUD measures among US adults at the general or sub-population level. Independent reviews were conducted by the authors to complete data synthesis and assess the risk of bias.

Results Overall, 46 studies validating substance use/SUD ($n = 46$) measures were included in this review, in which 63% were conducted in clinical settings and 89% assessed the validity of SUD measures. Among the studies that assessed SUD screening measures, 78% examined a generic SUD measure, and the rest screened for specific disorders. Almost every study used a different survey measure. Overall, sensitivity and specificity tests were conducted in over a third of the studies for validation, and 10 studies used receiver operating characteristics curve.

Conclusion Findings suggest a lack of standardized methods in surveys measuring and reporting prevalence of substance use/SUD among US adults. It highlights a critical need to develop short measures for assessing SUD that do not require lengthy, time-consuming data collection that would be difficult to incorporate into population-based surveys assessing a multitude of health dimensions.

Systematic review registration PROSPERO CRD42022298280.

Keywords Systematic review, Validity testing, Substance use, Measures

*Correspondence:

Ruchi Bhandari

ruchi.bhandari@hsc.wvu.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Substance use remains a serious adverse health risk in the United States (US). Forty million Americans reported illicit drug use in the past month in 2021, among people aged 12 years or older (Substance Abuse and Mental Health Services Administration, 2022b), with over 106,000 people in the US fatally overdosing in 2021 (National Institute on Drug Abuse, 2023). This is a dramatic increase of approximately 15% in overdoses within 1 year, signifying critical, life-threatening substance use problems and an associated overdose epidemic throughout the country. Notably, substance use problems that met the criteria for a substance use disorder (SUD) were reported by a sizeable proportion of the US population. More than 46 million people aged 12 years or older met the *Diagnostic Statistical Manual of Mental Disorders* (DSM-V) criteria for SUD in the past year, according to the National Survey of Drug Use and Health (NSDUH), with the highest percentage of people with SUD being young adults aged 18–25 (25.6%), followed by adults aged 26 or older (16.1%) [1]. Unfortunately, population-based assessments for SUD are rare beyond the NSDUH, especially at substate levels, although imperative to inform appropriate resource allocation and population-based interventions for states responding to the SUD and overdose epidemics.

There are few population-based surveys conducted in the US that assess substance use and/or SUD. NSDUH is a good example of a survey that monitors annual national trends in substance use and mental health issues in the US and provides estimates of the need for substance use prevention and treatment programs [2]. However, it involves lengthy questions and branching logic that are not feasible for use in surveys covering multiple health domains. Another validated tool to assess SUD is the National Addictions Vigilance Intervention and Prevention Program (NAVIPPRO™) Addiction Severity Index-Multimedia Version® (ASI-MV®) [3]. However, results from this measure may not be generalizable because it is only used to evaluate those already seeking SUD treatment. In addition, selection bias is likely because the participants are selected based on convenience sampling among treatment centers [4]. Other measures that have been validated for assessing substance use in the US are Drug Abuse Screening Test (DAST) [5], Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST), and tobacco, alcohol, prescription medication, and other substance use (TAPS) [6]. However, these survey measures also require multiple, lengthy questions to estimate the prevalence of SUD.

Validated substance use and SUD measures that are shorter and more versatile are needed to ease the incorporation of these measures into more multidimensional

population health surveys to better assess and respond to the current US substance use and overdose epidemics. Much work has been done on validating alcohol and tobacco measures, such as ASSIST and TAPS [6]. We know of no review of validation research conducted on other substance use and/or SUD measures among the US population, although previous studies provide valuable insights into measures assessing the efficacy of substance use measures and interventions [7] and addressing psychometric properties of screening tools among specific settings or populations [8]. Thus, the purpose of this review is to comprehensively summarize published literature investigating the validity of substance use and SUD measures, other than alcohol and tobacco use, in US surveys to advance the use of these validated measures on more population-based surveys.

Methods

Search strategy

This systematic review has followed the Preferred Reporting Items for Systematic reviews and Meta-analyses (PRISMA) guidelines [9] and was registered through PROSPERO (CRD42022298280). Potential eligible studies were identified by using the following nine electronic databases, starting from their inception up to November 22, 2021: PubMed, Scopus, CINAHL, PsycINFO, Academic Search Complete, Web of Science, ProQuest Theses and Dissertation Global, and Google Scholar. Primary keywords and phrases used for searching included “healthcare survey,” “mental health,” “substance use,” and “validity.” Detailed search strategies corresponding to the specific databases are shown in Supplementary Table 1.

The following study inclusion criteria were established a priori for use in this systematic review: [1] Utilized existing surveys or questionnaires at the county level or higher (validation may have been done at a sub-population level) or at clinical settings in the US; [2] to ensure the reviewed measures are applicable to US populations, and only studies conducted in the US were included in this review; [3] validity/validation testing conducted for measures of mental health and/or substance use; [4] study sample consisted of adults 18 years of age or older; [5] studies published in English language; and [6] peer-reviewed, published studies, official reports from surveys, and doctoral dissertations. In addition, exclusion criteria were applied to those studies that [1] assessed the validity of measures unrelated to mental health/substance use (i.e., physical activity, chronic disease, infectious disease); [2] assessed the validity of alcohol and/or tobacco measures only; [3] were published as abstract only or did not have full texts available; [4] were protocols, editorials, reviews, or commentary; [5] validated language translation or cultural version of an instrument; and [6] were

conducted internationally. In order to better align with the aims of this review, studies validating only alcohol/tobacco use measures were excluded because they have been widely studied in previous literature [10–13].

Quality assessment

An adapted risk-of-bias tool was developed for the purpose of this systematic review to assess the validity of substance use and mental health survey instruments. This methodological quality assessment tool was adapted from a previously published tool which evaluated the rigor of validity testing in the Behavioral Risk Factor Surveillance System (BRFSS) literature [14]. The new risk-of-bias tool was used to assess the quality of the [1] methodology and [2] statistical analyses of studies included in the systematic review. The methodological component was scored from 0 to 3 (3=studies utilizing a physical measurement(s) as a comparator during validity testing, which were considered to be the “gold standard,” 2=studies using measures other than actual physical measure, 1=studies that conducted face validity based on the researcher’s judgment or a collective judgment, 0=studies that did not report on the measurement used for validity testing). The statistical analysis component was scored from 0 to 2 (2=using statistical analyses such as sensitivity and specificity, correlation coefficient, or mean difference, 1=reporting prevalence estimates only, 0=no information on statistical analysis was reported). The methodological and statistical component scores were then totaled for an overall quality assessment score. Total scores ranged from 0 to 5, with 5 demonstrating the highest quality.

Data synthesis

All identified studies were imported to an EndNote library. After removing duplicates, the initial title and abstract screenings were conducted independently by three reviewers (Y. T., N. W., E. O.) using the pre-established inclusion and exclusion criteria. It was followed by the full-text review conducted independently by three reviewers (Y. T., E. C., R. M.) for the first 10% of the included studies. They then convened to review their selections to ensure agreement and refine criteria. Interrater reliability was calculated in STATA [15] using the Gwet’s AC to ensure agreement [16]. The remaining 90% of the selected articles were then split between the three reviewers for full-text review. Articles where a reviewer was not sure if they should be included or excluded were discussed among the three reviewers and decided by the senior author for final selection.

A data extraction form was created in Microsoft Excel to facilitate data extraction and synthesis. The form could capture up to 46 variables for each study. These

variables were grouped into four main categories: study characteristics (authors, reference, year of publication, and name of journal), measure characteristics (whether the measure was used for disorder screening, the SUD being assessed by the measure, response rate, study duration, items measured, recall period, and recruitment procedure), participant characteristics (overall health status, age, sex, race, income, education), and validation methods (type of validation, statistical analysis, comparison measure, and key results). Additionally, a single article could be considered as multiple studies if it validated measures among multiple study populations. Articles that validated multiple survey measures among the same study population were considered to be one study. We evaluated the different types of validity using pre-established definitions to standardize the understanding of validity among reviewers. Our focus was on examining criterion validity (including concurrent, predictive, and content validity) and construct validity (encompassing convergent, discriminant, and factorial validity). Specifically, criterion validity was examined through comparisons with “gold standard” measures where available or through the use of clinically established diagnostic criteria and outcomes. Face validity was determined if the article could demonstrate the extent to which a substance use measured what it intended to measure. Lastly, construct validity was assessed through statistical analyses examining the correlation between survey measures and related constructs, thus ensuring that measures accurately reflect the theoretical components of substance use and SUDs. Articles that did not specify the validation methods were discussed among the three reviewers and decided by the senior author for consensus if discrepancies existed.

All data were coded independently by two reviewers (Y. T., E. C.). After extracting data from the first 10 articles, the two reviewers met to discuss any discrepancies among coding strategies. Disagreements were brought to the senior author (R. B.) for conflict resolution. Although the inclusion and exclusion criteria were determined a priori, the completion of data extraction demonstrated unique differences present between mental health and substance use studies that evaluated the psychometric properties of their respective measures. As the study developed, the results gathered from the data synthesis for substance use were substantially different from mental health assessment, and the authors determined that these separate domains would be better discussed in two separate manuscripts. Thus, the results presented in this study are from studies that validated substance use measures identified in our search.

Results

Study characteristics

A total of 6950 results were initially obtained from the search. An additional 153 articles were identified by reviewing BRFSS reference lists [17]. A flow diagram documenting the search process and reasons for excluding studies is shown in Fig. 1. Of the 7103 articles, 2339 were duplicates and were excluded before the abstract/

title review. After reviewing 4764 abstracts/titles, 3744 articles were excluded. Of the 1020 articles, a full-text review of the first 10% of articles demonstrated an almost perfect inter-rater reliability agreement between reviewers on which articles met the inclusion criteria (Gwet’s AC: 0.8517 (0.8000–1.0000)). Following review of the full article text, 899 articles were removed. The key reasons for excluding the articles were because they [1] did not

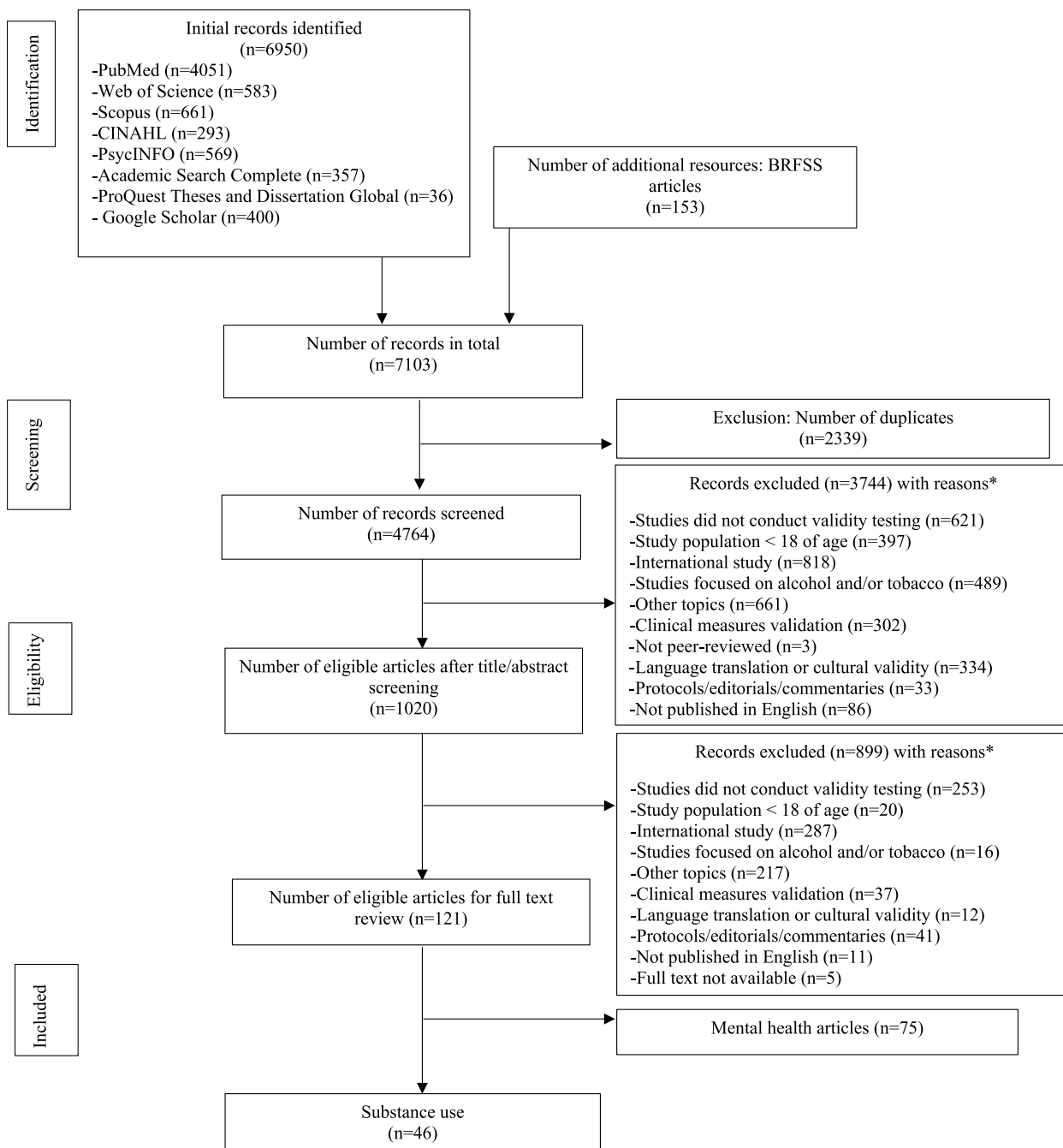


Fig. 1 Flow chart for the selection of studies*. *Studies could have been excluded for multiple reasons

conduct validity testing ($n=874$), [2] were conducted outside the United States ($n=1105$), or [3] were focused on topics other than substance use ($n=878$). For this review, a total of 46 articles met the inclusion criteria (Fig. 1). The characteristics of those 46 selected studies are presented in Table 1.

The included studies were published between 1979 and 2021, with a wide variation in demographic characteristics. Of the 46 studies, seven had over 80% male participants (Han et al., 2017; Peters et al., 2000; Tiet et al., 2016; Tiet et al., 2019; Tiet et al., 2015; Tiet et al., 2017; Zanis et al., 1994). Among these studies, two recruited only male participants (Peters et al., 2000; Zanis et al., 1994). Additionally, there was one study that only recruited female participants [22]. Racial and ethnic differences also existed among these study samples. Six studies had study sample of primarily (70% or more) White participants [41, 58], and three studies had a study population sample of 70% or more Black/African American (AA) participants [32, 42, 58]. Furthermore, 12 studies only recruited White and Black/AA participants [6, 18, 21, 24, 25, 31, 32, 36, 39, 41, 58, 59]. Seven studies (15.2%) did not report information on race/ethnicity characteristics [19, 22, 27, 34, 55, 57].

All 44 studies included in this review reported the final sample size, with a mean of 1427 (median=449) participants with an overall range of 23–10,167 participants. Only 13 studies reported response rate, and the response rates ranged between 13.4 [18] and 100% [32]. Twenty-six studies reported the survey duration, and it ranged from 1 month [32] to 120 months [20], with mean 28.48 months (median 13 months). Moreover, studies reported the mean age of the participants as <30 years ($n=4$), between 30 and 39 years ($n=16$), and ≥ 40 years ($n=18$). Another eight studies reported age groups or median age of the study population. Additionally, a majority ($n=37$) of the studies were conducted in non-population-based clinical settings (e.g., inpatient, outpatient).

Participant recruitment strategy

The participant recruitment strategies from included studies in this review were shown in Table 2. Of the 46 studies, only 4% ($n=2$) examined SUD in the general population [20, 28]; the rest ($n=44$) of the studies were conducted in clinical or other population subgroups. In the first population-based study, 6664 adult Medicaid enrollees were recruited from 1 of 7 Florida regions who took part in the Florida Health Services Survey at least once between 1998 and 2008 [20]. Researchers assessed the internal psychometric properties of the Simple Screening Instrument for Substance Abuse (SSI-SA) but did not compare survey responses with SUD diagnoses in Medicaid clinical records. In the second

population-based study, participants were selected from the National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III) sample, which included noninstitutionalized US adult residents (aged 18 years or older) [28]. The authors then selected 777 respondents for the procedural validity study and used a test–retest design to compare concordance of respondents' answers to the NESARC-III survey questions with a semi-structured interview, the Psychiatric Research Interview for Substance and Mental Disorders, DSM-5 version (PRISM-5), administered by a clinician.

Of the remaining 44 studies not in the general population, over three quarters ($n=35$) were conducted in the clinical setting, with the majority ($n=19$) in the inpatient setting [6, 19, 21, 22, 24, 26, 32, 35–39, 42, 46, 48–50, 59]. Eleven studies were conducted in the outpatient clinical setting [5, 23, 31, 40, 43, 47, 51–55], 5 studies were conducted in both the inpatient and outpatient settings (Harris et al., 2015; Hser et al., 1999; Kellogg et al., 2002; Kupetz et al., 1979; Salyers et al., 2000), and 1 study was conducted in a Veterans' Administration shelter [58]. The remaining studies ($n=8$) were conducted outside the clinical setting. For example, participants were recruited from an alcohol and drug program [18], prison substance abuse treatment programs [25], Holiday Transfer Facility [41], and a novel jail-release program [56]. Lastly, four studies consisted of sub-population samples within the National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III) [29] and universities using student participants [45, 57].

Quality of studies

Risk of bias was assessed based upon the methodology used for instrument comparison and the statistical analysis conducted. Although several studies adopted recruitment strategies that limited their study population to specific groups (for example, only recruiting male or white populations), the risk-of-bias assessment employed by the current study did not account for recruitment. As a result, most of the included studies ($n=41$) had a risk-of-bias score of 4 or higher (Table 1). Two studies had a score of 3 [30, 34], one studies had a score of 2 [29, 48], and two studies had a score of only 1 [27, 44]. Among those studies with low-quality assessment scores, four studies lacked statistical comparisons and reported prevalence estimates only [27, 30, 34, 44]. There were three studies that did not report on validation methodology [27, 44, 48].

Survey measure

Among the articles included in this review, 89% ($n=41$) used measures specifically designed for screening SUDs. For example, seven studies tested the validity

Table 1 Characteristics of included studies of validation testing

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Alexander and Leung [18]	<ul style="list-style-type: none"> • Aged ranged from 18 to 54 with mean age of 33.2 • Most of participants were male ($n=64$, 58.7%) • 63% White, 20% Black, and 17% Hispanic 	Non-population-based study evaluated the concurrent, convergent, and discriminant validity between the MSI-X and five other instruments in the alcohol and drug program	MSI-X <ul style="list-style-type: none"> • Specific substance use: Marijuana use • Sample size: $n = 107$ • Response rate = 13.4% • Survey duration: September 2001–June 2002 	Construct (convergent) validity <ul style="list-style-type: none"> • Comparison measure: Compared with SASSI-3 overall 'decision rule' for those most likely to be diagnosed with substance dependence or substance abuse • Statistical analysis: Two separate independent sample t-tests Criterion (concurrent) validity <ul style="list-style-type: none"> • Comparison measure: Comparing with DSM-G-MI, DAST-20, SASSI-3, and ASI interviewer severity rating scales and selected variables • Statistical analysis: Correlation coefficients 	Those who had a high probability of substance dependence had a significantly higher number of MSI-X problems (mean = 6.2) than those who had a low probability for substance dependency (mean = 2.4). The concurrent validity analysis reveals excellent, strong significant positive association between MSI-X scores and the DSM-G-MI ($r=0.811$; $p < .001$), good to moderately strong significant positive associations with the DAST-20 scores ($r=0.531$; $p < .001$), and two specific SASSI-3 subscale scores, the Face Valid Other Drugs (FVOD) subscale scores ($r=0.626$; $p < .001$) and correlational (COR) subscale ($r=0.522$; $p < .001$)	5
Appleby, Dyson [19]	<ul style="list-style-type: none"> • Mean age of 32 years 	Non-population-based study conducted in the inpatient clinical setting and used CUAD by comparing CUAD severity ratings with severity ratings from the other instruments used in the study and by determining its sensitivity and specificity in relation to SCID diagnoses of substance use disorders	CUAD <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: 100 • Response rate: Not reported • Survey duration: Not reported 	Construct validity <ul style="list-style-type: none"> • Comparison measure: Comparing CUAD severity ratings with severity ratings from other instruments used in the study: SMAST, DAST, ASI • Statistical analysis: Chi-square statistical analysis Criterion validity (concurrent validity) <ul style="list-style-type: none"> • Comparison measure: Comparing CUAD severity ratings with severity ratings from ASI alcohol and drug scales • Statistical analysis: Sensitivity 	This study showed a highly significant positive correlation between the CUAD alcohol subscale and the SMAST and an even stronger association between the CUAD drug subscale and the DAST. Additionally, high positive correlations between CUAD and ASI alcohol and drug scales and both current substance use measures provided additional evidence of concurrent validity. Seventy-three patients had a SCID lifetime diagnosis of an alcohol use disorder, and the CUAD correctly identified 51 of them (70% sensitivity), which also showed evidence criterion validity	4

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Boothroyd, Peters [20]	<ul style="list-style-type: none"> • Mean age of 40.1 ($SD = 12.2$) • 25.4% were males, and 74.6% were females 	Population-based study examined the SSI's psychometric properties within Medicaid recipients in Florida	SSI-SA <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 6664$ • Response rate: 35–45% • Survey duration: 1998–2008 	Construct validity (convergent validity) <ul style="list-style-type: none"> • Comparison measures: Examine the association of respondents' SSI-SA classification at cutoff of 4 vs. at or above the cutoff 4 with their scores on a measure of general functioning, an indicator of current substance use, and quality of life • Statistical analysis: Independent t-tests sensitivity, specificity, PPV, NPV, ROC curve analysis 	The SSI-SA had excellent internal consistency (0.85). Evidence of the SSI's validity was strong. Using the recommended SSI-SA cutoff score of 4 or higher to indicate the presence of a substance abuse problem, the SSI-SA had respectable sensitivity (0.82) and specificity (0.90)	4
Broderick, Richmond [21]	<ul style="list-style-type: none"> • Mean age (SD) = 43 (15) years • Most of participants were males (57% vs. 43%) • Hispanic (37%), White (41%), and Black (18%) 	Non-population-based study conducted in the inpatient clinical setting to evaluate two brief screen questions to assess the degree to which these single-item screening questions detected risky substance use compared to a longer, validated screening tool	Two-item brief screen (survey name not reported) <ul style="list-style-type: none"> • Specific substance use: Illicit drugs and marijuana • Sample size: $n = 1218$ • Response rate: $n = 72\%$ • Survey duration: August 25–October 31, 2010 	Construct validity <ul style="list-style-type: none"> • Comparison measure: ASSIST • Statistical analysis: Sensitivity and specificity 	Sensitivity values for the marijuana and street drug questions were 72% and 40% respectively. Specificity values for the marijuana and street drug questions were 96% and 99%, respectively	5
Carter, Yu [6]	<ul style="list-style-type: none"> • Mean age (SD) = 49 (14.9) • Most participants were female (61.9%) • White (92.5%), Black/African American (4.5%), and Hispanic (0.9%) 	Non-population-based study conducted in the community inpatient clinical setting to validate the performance of TAPS tool compared to a reference-standard substance use assessment	TAPS <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 1523$ • Response rate: Not reported • Survey duration: November 2019 to October 2020 	Construct validity <ul style="list-style-type: none"> • Comparison measure: ASSIST • Statistical analysis: AUC of ROC curve 	The TAPS tool showed fair or better discrimination between moderate risk use and high-risk use for tobacco, alcohol, and prescription opioids ($AUCs: 0.75–0.97$) and fair or better discrimination between low-risk and moderate-risk use in five of eight subscales, including tobacco, alcohol, marijuana, stimulants, and heroin ($AUCs: 0.70–0.92$)	5

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Chasnoff, Wells [22]	All pregnant women 18 years of age or older	Non-population-based study conducted in the inpatient clinical setting to validate the 4P's Plus to identify women whose substance-use levels fall below the DSM-IV criteria but still at risk from any level of use of alcohol or illicit drugs	4P's Plus <ul style="list-style-type: none"> Specific substance use: Not reported Sample size: $n = 228$ Response rate: 59% Survey duration: Not reported 	Criterion validity (predictive validity) <ul style="list-style-type: none"> Comparison measure: Comparing the 4P's Plus positive and negative screens with positive and negative clinical assessment Statistical analysis: Sensitivity, specificity, PPV, NPV 	The overall reliability for the 5-item measure was 0.62. Sensitivity and specificity were very good at 87 and 76%, respectively. Positive predictive validity was low (36%), but negative predictive validity was quite high (97%). Of the 31 women who had a positive clinical assessment, 45% were using less than 1 day per week	5
Dennis and Davis [23]	<ul style="list-style-type: none"> Mean age (SD) = 31 (11) years 59% of the study population were males, and 41% were females White: 64%, African American/Black: 10%, and Other race: 10% 	Non-population-based study conducted in the outpatient clinical setting to examine the psychometric properties of core GAIN-Q3 assessment	GAIN-Q3 <ul style="list-style-type: none"> Specific substance use: Not reported Sample size: $n = 10,167$ Response rate: Not reported Survey duration: 2002–2010 	Construct validity (convergent and discriminant validity) <ul style="list-style-type: none"> Comparison measure: Compared the correlation between the screener and the full GAIN-I scale Statistical analysis: Pearson correlation matrix between the shortened screener and the full-length scores, ROC analysis 	Despite the condensed lengths of the screening measures compared with their longer versions, the reliability estimates are within the good to excellent range (0.7 to 0.9) in terms of internal consistency for 7 of the 10 screeners for adults. Moreover, there is strong evidence for the measures' convergent and discriminant validity and efficiency (i.e., maximum information gathered in as few items possible) relative to the full-length scales as well as relative to other scales in the full GAIN-I	5
Dezman, Gorelick [24]	<ul style="list-style-type: none"> All participants aged over 18 years old Most participants were male ($n = 803$; 72.0%) Caucasian: $n = 627$ (56.2%), African American: $n = 481$ (43.1%), Hispanic: $n = 7$ (0.6%) 	Non-population-based study conducted in the trauma inpatient clinical setting to test characteristics of a 4-item drug CAGE questionnaire to detect DUDs	CAGE <ul style="list-style-type: none"> Specific substance use: Not reported Sample size: $n = 1115$ Response rate: 58% Survey duration: September 1994–November 1996 	Criterion validity <ul style="list-style-type: none"> Comparison measure: Comparison with SCID-generated DUD diagnoses as the standard Statistical analysis: Sensitivity, specificity, PPV, NPV, and AUC 	The drug CAGE screen had an AUC = 0.9. Each individual question had a high AUC (0.78–0.87). The drug CAGE questionnaire had high AUCs across all sociodemographic and injury mechanism subgroups, both for each individual question and overall	5

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Duncan, Sacks [25]	<p>The study showed self-reported demographic characteristics of the study sample composed of African American, Latinos, and Whites. The majority of the sample in each racial and ethnic group was male. The average age was 36.03 years for African Americans, 32.74 years for Latinos, and 36.02 years for Whites. Males: 56.03% for African Americans, 60.8% for Latinos, and 36.02% for Whites</p>	<p>Non-population-based study conducted in the prison substance abuse treatment programs to test the stability of the performance of the CODSI-MD and SMD across three racial/ethnic groups of offenders entering prison substance abuse treatment programs</p>	<p>CODSI-MD</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: n = 353 • Response rate: Not reported • Survey duration: Not reported <p>CODSI-SMD</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: n = 353 • Response rate: Not reported • Survey duration: Not reported 	<p>Criterion validity</p> <ul style="list-style-type: none"> • Comparison measure: Compare three racial or ethnic groups of offenders entering prison substance abuse treatment programs • Statistical analysis: Sensitivity, specificity, PPV, NPV, positive and negative likelihood ratio <p>Criterion validity</p> <ul style="list-style-type: none"> • Comparison measure: Compare three racial or ethnic groups of offenders entering prison substance abuse treatment programs • Statistical analysis: Sensitivity, specificity, PPV, NPV, positive and negative likelihood ratio 	<p>No statistical differences in sensitivity or specificity for either CODSI-MD or SMD across the African American, Latino, and White prisoner groups</p>	4
Han, Sherman [26]	<ul style="list-style-type: none"> • Mean age (SD) = 47.1 (13.4) years • Most participants were males (72.0%) • White: 56.2%, African American/Black: 40.8%, American Indian/Alaska Native: 1.8%, Asian: 2.5%, Other race: 22.1% 	<p>Non-population-based study conducted in the inpatient clinical setting to assess the diagnostic accuracy of the SUBS in comparison to ASSIST and AUDIT-C to identify unhealthy and high-risk alcohol and drug use among hospitalized current smokers</p>	<p>SUBS</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: n = 439 • Response rate: Not reported • Survey duration: Not reported 	<p>Criterion validity (concurrent)</p> <ul style="list-style-type: none"> • Comparison measure: AUDIT-C for alcohol use and ASSIST for drug use • Statistical analysis: ROC analysis, sensitivity, specificity 	<p>The SUBS had a sensitivity of 98% (95% CI 95–100%) and specificity of 61% (95% CI 55–67%) for unhealthy alcohol use, a sensitivity of 85% (95% CI 80–90%) and specificity of 75% (95% CI 78–87%) for illicit drug use, and a sensitivity of 73% (95% CI 61–83%) and specificity of 83% (95% CI 78–87%) for prescription drug nonmedical use. For identifying high-risk use, a higher cutoff (response of “3 or more days” of use indicates a positive screen), the SUBS retained high sensitivity (77–90%), and specificity was 62–88%</p>	4
Harris, Ellerbe [27]	<p>Not reported</p>	<p>Non-population-based study conducted in the inpatient and outpatient clinical setting to assess the specification validity of the 2009 HEDIS substance use disorder initiation and engagement</p>	<p>HEDIS</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: n = 2726 • Response rate: Not reported • Survey duration: Not reported 	<p>Criterion validity (specification validity)</p> <ul style="list-style-type: none"> • Comparison measure: Not reported • Statistical analysis: PPV 	<p>The PPV were excellent (> 90%) for residential and outpatient records selected from addiction treatment programs but more modest for records generated in non-addiction settings and were highly variable across facilities</p>	1

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Hasin, Greenstein [28]	<ul style="list-style-type: none"> • More participants were females than males (54.5% vs. 45.5%) • White: 67.8%, African American/Black: 22.5%, American Indian/Alaska Native: 2.1%, Asian: 0.7%, and Other race: 6.9% 	<p>Population-based study used a test-retest design to compare concordance of the NESARC-III survey questions with a semi-structured interview, the Psychiatric Research Interview for Substance and Mental Disorders, DSM-5 version (PRISM-5) administered by a clinician</p>	<p>AUDADIS</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 712$ • Response rate: 92.5% • Survey duration: June 2012–July 2013 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: Physician diagnosis • Statistical analysis: Intraclass correlation coefficients 	<p>Concordance of the AUDADIS-5 and the PRISM-5 for DSM-5 diagnoses of substance use disorders ranged from fair to good ($\kappa = 0.40–0.72$). Concordance on dimensional scales was excellent ($ICC \geq 0.75$) for the majority of DSM-5 SUD diagnoses and fair to good ($ICC = 0.43–0.72$) for most of the rest</p>	5
Hasin, Keyes [29]	<p>For frequent marijuana users: 66.9% males and 33.1 females, 75.4% White, 11.4% African American/Black, 3.7% American Indian/Alaska Native, 2.2% Native Hawaiian or Other Pacific Islander</p> <p>For marijuana use only users: 58.2% males and 41.8% females, 71.1% White, 15.3% African American/Black, 2.7% American Indian/Alaska Native, 2.7% Native Hawaiian or Other Pacific Islander</p>	<p>Sub-population-based study and participants were selected from the NESARC-III sample to investigate the factor structure, clinical validity, and psychiatric correlates</p>	<p>NESARC</p> <ul style="list-style-type: none"> • Specific substance use: Marijuana • Sample size: $n = 3732$ • Response rate: $n = 81\%$ • Survey duration: 2001–2002 	<p>Construct validity (convergent validity)</p> <ul style="list-style-type: none"> • Comparison measure: AUDASIS-IV • Statistical analysis: Factor analysis <p>Criterion validity (predictive validity)</p> <ul style="list-style-type: none"> • Comparison measure: DSM-IV marijuana dependence criteria • Statistical analysis: Binomial regression 	<p>Both marijuana withdrawal symptoms were associated with significant distress/impairment, substance use to relieve/avoid marijuana withdrawal symptoms, and quantity of marijuana use. Panic and personality disorders were associated with anxiety symptoms in both frequent marijuana users and marijuana-only users</p>	4
Hser, Shen [30]	<ul style="list-style-type: none"> • The overall mean age was 32.6 years ($SD = 7.6$) • Two-thirds (66.0%) of the DATOS sample were male • Almost half of the patients (46.6%) were African American, 38.3% were white, 12.5% were Hispanic, and the remainder (2.7%) were either Asian or others 	<p>Non-population-based study conducted in the inpatient and outpatient clinical setting to develop a lifetime severity index for cocaine use disorder and examine its predictive validity of posttreatment outcome using data from the nation Drug Abuse Treatment Outcome Study</p>	<p>LSI-Cocaine</p> <ul style="list-style-type: none"> • Specific substance use: Cocaine use disorder • Sample size: $n = 2107$ • Response rate: Not reported • Survey duration: 1993–1999 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: Not reported • Statistical analysis: Concordance agreement 	<p>A higher value of the index, indicating greater severity, predicted a greater likelihood of relapse (the odds ratios were 5.7 for high severity and 4.4 for medium severity, relative to low severity) and shorter time to relapse. Similarly, the polytomous logistic analysis indicated that the index predicted levels of posttreatment cocaine use (odds ratios of daily use were 47.8 for the high severity and 18.8 for medium severity; the corresponding odds ratios of weekly use were 6.75 and 5.10 and for less than weekly use were 3.35 and 3.57)</p>	3

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Jackson, Covell [31]	<ul style="list-style-type: none"> • Mean age (SD) = 37 (7.8) • 71% males and 29% females • 41% were White, and 59% were African American/Black 	<p>Non-population-based study conducted in the outpatient clinical setting to clarify the concordance rates of self-report of drug use compared to results of urine screens for drugs of abuse among individuals diagnosed with both a mental health and a substance use disorder</p>	<p>Self-reported drug use reported</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 196$ • Response rate: Not reported • Survey duration: 1992–1998 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: Urine drug screen • Statistical analysis: Concordance agreement (concordance between self-reported and urine drug screen) 	<p>The concordance between self-report and results from urine screens was high. Estimates for the likelihood of use of marijuana and cocaine within the past 30 days were 15% and 32%, respectively, based on urine screens, 25% and 35% based on self-report, and 28% and 43% based on information from both sources combined</p>	5
Joyner, Wright [32]	<ul style="list-style-type: none"> • Mean age = 30 years • 70% males and 30% females • 22% were White; 78% were African American/Black 	<p>Non-population-based study conducted in inpatient clinical setting to examine the ability of the ASI to provide valid and reliable data within a homeless population of drug misusers</p>	<p>ASI reported</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 23$ • Response rate: 100% • Survey duration: May–June 1991 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: Correlation between ASI section and corresponding composite measure • Statistical analysis: Pearson product-moment correlation coefficients 	<p>Only three of the coefficients did not meet this criterion. Years of education and employment income were more highly correlated with composites outside the area of employment. The legal composite rating was more highly correlated with days of cocaine/crack use rather than length of last incarceration. The mean composite score in the study sample was 0.102 for those with no alcohol problem, 0.403 for those with alcohol as a secondary problem, and 0.694 for those identifying alcohol as their primary problem</p>	4
Kellogg, Ho [33]	<p>All drug positive participants: Mean age (SD) = 40.0 (9.8), most participants were male (64%), Caucasian (38%), African American (29%), Hispanic (26%), Other/mixed race (7%), mean education year (SD) = 12.5 (2.5)</p> <p>Normal volunteers: Mean age (SD) = 35.7 (14.1), male participants 49%, Caucasian (56%), African American (16%), Hispanic (14%), Other/mixed race (1%), Asian (13%), mean education year (SD) = 15.7 (3.2)</p>	<p>Non-population-based study conducted in the inpatient and outpatient clinical setting to assess the validity of the DRG scale through multiple analyses</p>	<p>PAI DRG reported</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 370$ • Response rate: Not reported • Survey duration: Not reported 	<p>Construct validity</p> <ul style="list-style-type: none"> • Comparison measure: Compared with ASI drug composite scores and ASI drug severity ratings • Statistical analysis: Kruskal–Wallis test • Content validity • Comparison measure: Not reported • Statistical analysis: Sensitivity and specificity 	<p>The Kruskal–Wallis tests revealed that there were significant correlations among the PAI DRG scale and the ASI scales related to the frequency of use, negative consequences of use, and need and desire for treatment</p>	5

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Kupetz, Klagsbrun [34]	Not reported	Non-population-based study conducted in the inpatient and outpatient clinical setting to develop a self-assessment survey instrument for the purpose of detecting abuse of both alcohol and other drugs and to assess validity	SUAS <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 5745$ • Response rate: 88% • Survey duration: Not reported 	Construct validity (convergent validity) <ul style="list-style-type: none"> • Comparison measure: Medical chart • Statistical analysis: Prevalence, prevalence percent agreement, detection 	The SUAS was far superior to patient charts in identifying heavy drug and/or alcohol use, especially when both quantity-frequency and social consequences criteria were considered. Also, the SUAS did at least in simply identifying current heavy use of substances, although the numbers are too small to permit definite conclusions. However, an internal validity, which is the correlation of various parts of the SUAS, was not included in the present study	3
Leonhard, Mulvey [35]	<ul style="list-style-type: none"> • Mean age (SD) = 34.46 (9.82) • More participants were males (72.2% vs. female 27.1%) • White (47.4%), African American/Black (38.1%), American Indian/Alaska Native (0.7%), Asian (0.5%), Other race (11.3%) 	Non-population-based study conducted in a busy online inpatient alcohol and drug abuse clinical setting without tight procedural controls to investigate the internal consistency and validity	ASI <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 8984$ • Response rate: 90% • Survey duration: Not reported 	Construct validity (convergent and discriminant validity) <ul style="list-style-type: none"> • Comparison measure: Interviewer severity ratings and composite scores • Statistical analysis: Pearson correlation coefficients 	Validity analyses showed good promise. Correlation matrices for both composite scores and severity ratings demonstrated good evidence for discriminant validity in measuring the ASI's seven dimensions, medical severity, employment problems, alcohol and drug use, legal problems, social difficulties, and psychiatric problems	4

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
McGovern and Morrison [36]*	<ul style="list-style-type: none"> • Mean age = 36 years, male: 67% and female: 33% • Most participants were White (50%), followed by Black (42%) and other (1%) 	<p>Non-population-based study conducted in the inpatient clinical setting to examine criterion-related validity by comparing the CUAD's concordance rates with psychiatrist-derived diagnoses</p>	<p>CUAD</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 129$ • Response rate: Not reported • Survey duration: Not reported 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: Comparing between CUAD with the MAST and DAST instruments • Statistical analysis: Comparison values in parentheses (%) 	<p>The CUAD total severity score is positively associated with all three criterion measures at highly significant levels. The CUAD total severity score is able to significantly discriminate the level of care assignments ($p < .001$). The validity of the CUAD is reported and appear satisfactory</p>	4
	<ul style="list-style-type: none"> • Mean age = 35.8 years • Most participants are male ($n = 259$, 74%) • Most participants were White (70%), followed by Black (25%) and Hispanic (5%) 	<p>Non-population-based study conducted in the inpatient clinical setting to further investigate the validity of the CUAD by comparing two frequently used measures, MAST and DAST</p>	<p>CUAD</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 348$ • Response rate: Not reported • Survey duration: Not reported 	<p>Construct validity</p> <ul style="list-style-type: none"> • Comparison measure: The comparison was made by contrasting the CUAD derived DSM-III-R substance use disorder diagnoses with the chart diagnosis determined by the unit psychiatrists • Statistical analysis: Pearson correlation coefficients <p>Criterion validity (concurrent validity and predictive validity)</p> <ul style="list-style-type: none"> • Comparison measure: Explored by testing the capacity of the CUAD total severity score to distinguish among patients assigned to three levels of substance abuse treatment: inpatient, partial hospitalization, and outpatient • Statistical analysis: Relating the instrument concurrently with subjects scored on the MAST and DAST and predictively with the assignment of patients to varying levels of care 	<p>The CUAD total severity score is positively associated with all three criterion measures at highly significant levels. The CUAD total severity score is able to significantly discriminate the level of care assignments ($p < .001$). The validity of the CUAD is reported and appear satisfactory</p>	4

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
McNeely, Cleland [37]	<ul style="list-style-type: none"> Age ranges from 21 to 65, mean age (SD) = 46 (12), median age = 48 More participants were female ($n = 236$, 51.4%), followed by male ($n = 211$, 48.1%) More participants were Black/African American ($n = 238$, 51.0%), followed by Hispanic ($n = 93$, 20.2%), White ($n = 88$, 19.1%), Other ($n = 38$, 8.2%) Most of participants had education on some college or trade school ($n = 116$, 25.3%) 	<p>Non-population-based study conducted in the inpatient clinical setting to test the sensitivity and specificity of SISQs for alcohol and other drug use, as well as their feasibility, for self-administration</p>	<p>SISOs</p> <ul style="list-style-type: none"> Specific substance use: Not reported Sample size: $n = 459$ Response rate: Not reported Survey duration: June–July 2012 (site A), November 2013–June 2013 (site B) 	<p>Construct validity (discriminant validity)</p> <ul style="list-style-type: none"> Comparison measure: Not reported Statistical analysis: Sensitivity, specificity, positive and negative diagnostic likelihood ratios, ROC curves 	<p>The SISQ drug had sensitivity of 71.3% (95% CI 62.4–79.1) and specificity of 94.3% (95% CI 91.3–96.6), $AUC = 0.83$ (95% CI 0.79–0.87), for detecting unhealthy drug use, and sensitivity of 85.1 (95% CI 75.0–92.3) and specificity of 88.6% (95% CI 85.0–91.6), $AUC = 0.87$ (95% CI 0.83–0.91), for drug use disorder</p>	4
McNeely, Strauss [38]	<ul style="list-style-type: none"> Age ranges from 21 to 65, mean age (SD) = 46 (11.8), median age = 49 Males and females were equally distributed ($n = 292$, 49.8%) More participants were Black/African American ($n = 293$, 50.2%), followed by Hispanic ($n = 127$, 21.7%), White ($n = 109$, 18.7%), Other ($n = 51$, 8.7%) Most of participants had HS grad or GED degree ($n = 199$, 34.0%) 	<p>Non-population-based study conducted in the adult inpatient clinical setting to evaluate the validity and test–retest reliability of the SUBS</p>	<p>SUBS</p> <ul style="list-style-type: none"> Specific substance use: Not reported Sample size: $n = 586$ Response rate: Not reported Survey duration: April 2011–April 2012 (site A); June–July 2012 (site B) 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> Comparison measure: Compared with participants with saliva drug testing, the reference standard measures were used in combination to identify unhealthy use and substance use disorder, for each of the four substance classes in the SUBS Statistical analysis: Pearson correlation coefficients 	<p>For unhealthy use of illicit or prescription drugs, sensitivity was 82.5% (95% CI 75.7 to 88.0) and specificity 91.1% (95% CI 87.9 to 93.6). Analyses of area under the receiver operating curve (AUC) indicated good discrimination ($AUC 0.74–0.97$) for all substance classes</p>	4
Miele, Carpenter [39]	<ul style="list-style-type: none"> Mean age (SD) = 35.6 (8.7) Most participants were males (62%), non-Hispanic origin White (61%), African American (26%), and 1.3% Hispanic 2.3% had not completed high school 	<p>Non-population-based study conducted in the inpatient clinical setting to investigate the test–retest reliability, inter-rater consistency, diagnostic concordance, and concurrent validity of the SDSS's ICD-10 dependence and harmful use scales</p>	<p>SDSS</p> <ul style="list-style-type: none"> Specific substance use: Marijuana use Sample size: $n = 180$ Response rate: Not reported Survey duration: Not reported 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> Comparison measure: Compared with ICD-10 dependence diagnosis Statistical analysis: Bivariate correlations between CUDITR composite scores and measures of the frequency of marijuana use per week and marijuana-related consequences 	<p>The ICD-10 dependence and harmful use severity scales were significantly associated with GAS scores for heroin. For cannabis, the 3 ICD-10 dependence scales and the frequency of harmful use symptoms scale were significantly associated with the number of days of cannabis use</p>	4

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
O'Hare, Cutler [40]	<ul style="list-style-type: none"> • Mean age was 44.5 years old • More participants were females (54.2%) than males (45.8%) • The majority of participants were White (90.7%) 	<p>Non-population-based study conducted in the outpatient clinical setting to examine preliminary validity and reliability of the SSPI and test the concurrent validity of the SRSa, QFI, and one-item frequency of marijuana use index by examining their intercorrelations</p>	<p>SSPI</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Specific mental health: Multidimensional psychosocial distress scale • Sample size: $n = 227$ • Response rate = 76.4% • Survey duration: Spring–summer of 1999 	<p>Content validity (factorial validity)</p> <ul style="list-style-type: none"> • Comparison measure: Not reported • Statistical analysis: Principal components analysis with varimax rotation <p>Construct validity (specific type not reported)</p> <ul style="list-style-type: none"> • Comparison measure: Self-related SRSa, a QFI, and a one-item index measuring the frequency of marijuana use • Statistical analysis: Coefficient correlations (Spearman's rho)-correlations among substance abuse indices 	<p>Correlations among the one-item self-reported substance abuse index (SRSa), QFI (average drinks per week), and frequency of marijuana use were moderate and significant. Correlations among SSPI subscales and these three substance abuse indices tended to be insignificant overall, or if significant, very low</p>	4
Peters, Greenbaum [41]	<ul style="list-style-type: none"> • Mean age (SD) = 32.6 (10.2) years • All the participants were males • White (32.7%), African American (44.9%), and Other race (22.5%) 	<p>Non-population-based study conducted among a sample of 400 male inmates in the Holiday Transfer Facility to evaluate the effectiveness of eight different substance abuse screening instruments included overall accuracy, PPV, and sensitivity</p>	<p>ASI, DAST, SASSI-2, SSI and TCUDS</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 400$ • Response rate = 75% • Survey duration: Feb–April 1996 	<p>Criterion validity (specific type not reported)</p> <ul style="list-style-type: none"> • Comparison measure: Not reported • Statistical analysis: Coefficient correlations (Spearman's rho): correlations with brief SSPI subscales <p>Construct validity (specific type not reported)</p> <ul style="list-style-type: none"> • Comparison measure: Not reported • Statistical analysis: Coefficient correlations (Spearman's rho): correlations with brief SSPI subscales 	<p>The TCUDS, SASSI-2, and SSI were examined with respect to their utility to identify either alcohol or drug disorders, while other screening instruments (including independent alcohol and drug screens from the ASI) were examined with respect to alcohol or drug disorders. The positive predictive value in detecting either alcohol or drug dependence was highest for the TCUDS, followed by the ADS/ASI drug, the SSI, and the SASSI-2. Sensitivity of the multipurpose instruments in detecting either alcohol or drug dependence was highest for the SSI, followed by the ADS/ASI drug, the SASSI-2, and the TCUDS</p>	5

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Ramsay, Abedi [42]	<ul style="list-style-type: none"> • Mean age (SD) = 23.5 (4.6) years • More participants were males (70%) • The majority of participants were African American (91.7%). Other race (5%) 	<p>Non-population-based study conducted in the inpatient clinical setting to describe the development and initial validation of two new, multidimensional measures of substance use, LSUR and LSUR-12, and to provide structured visual tools to aid recall and enhance the validity of the data</p>	<p>LSUR</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n=60$ • Response rate: Not reported • Survey duration: Not reported 	<p>Construct validity (hypothesis testing validity)</p> <ul style="list-style-type: none"> • Comparison measure: Comparing LSUR and LSUR-12 scores for each substance use and a set of key scores from the LSUR and LSUR-13 • Statistical analysis: Spearman correlations, independent samples Student's <i>t</i>-tests, and Mann–Whitney <i>U</i>-tests 	<p>Lifetime and past 12-week doses measured by the LSUR and LSUR-12 were highly correlated with the HONC and FTND scores (ranging from $\rho=0.796$ to $\rho=0.852$). The number of years of education completed was negatively correlated with the lifetime nicotine dose, as predicted ($\rho=-0.351$). The LSUR can be used to obtain total alcohol consumption for the 5 years preceding assessment, which had an even stronger correlation with the RETROSUB ($\rho=0.768$, $p<0.001$, $n=43$). The LSUR can also estimate the number of days of cannabis use in the 5 years preceding the assessment, and for this, a stronger correlation was noted with the number of days of drugs use measured by the RETROSUB ($\rho=0.726$, $p<0.001$, $n=41$)</p>	4
Rosenberg, Drake [43]	<ul style="list-style-type: none"> • Mean age (SD) = 38.03 (8.82) years • More participants were females (52.2% vs. 48.8% for males) • The majority of participants were White (98.4%). Other race (1.6%) 	<p>Non-population-based study conducted in the outpatient clinical setting to test the validity of DALI compared with clinician diagnosis and other screens for substance use disorders</p>	<p>DALI</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n=247$ • Response rate: Not reported • Survey duration: 1994–1996 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: Clinician diagnosis • Statistical analysis: ROC analysis 	<p>ROC curves showed that the DALI functioned significantly better than traditional instruments for both alcohol and drug use disorders</p>	5

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Salyers, Bosworth [44]	<ul style="list-style-type: none"> • Mean age (SD) = 42.3 (10.1) • More participants were females (64.8% vs. 35.2% for males) • White (47.1%), African American (44.6%), and Other race (5.5%) 	<p>Non-population-based study conducted in the inpatient and outpatient setting with several mental illness to examine the internal factor structure of the SF-12, test-retest reliability, and convergent and divergent validity by comparing SF-12 scores to other indexes of physical and mental health</p>	<p>SF-12</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 801$ • Response rate: Not reported • Survey duration: June 1997–December 1998 	<p>Construct validity (specific type of validity not reported)</p> <ul style="list-style-type: none"> • Comparison measure: Not reported in detail, only specifying hospitalization and substance use disorder were associated with MCS but not with PCS. Chronic health problems and doctor visits for physical health were more strongly related to PCS than to MCS. However, the correlations for physical health hospitalizations did not differ significantly between PCS and MCS. Mental health indexes of self-reported overall mental health, psychiatric hospitalization, and substance use disorder were more strongly related to MCS than to PCS 	<p>Each of the physical health indexes were significantly related to PCS as well as to MCS. However, psychiatric hospitalization and substance use disorder were associated with MCS but not with PCS. Chronic health problems and doctor visits for physical health were more strongly related to PCS than to MCS. However, the correlations for physical health hospitalizations did not differ significantly between PCS and MCS. Mental health indexes of self-reported overall mental health, psychiatric hospitalization, and substance use disorder were more strongly related to MCS than to PCS</p>	1
Schultz, Bassett [45]	<ul style="list-style-type: none"> • Mean age (SD) = 20.03 (1.51) • Most of participants were females (70.1%) and White (87.8%) 	<p>Sub-population-based study conducted in a large, public southeastern university to evaluate the internal consistency, concurrent and discriminant validity, and item performance of CUDIT-R among college students who reported recent marijuana use</p>	<p>CUDIT-R</p> <ul style="list-style-type: none"> • Specific substance use: Marijuana use • Sample size: $n = 229$ • Response rate: Not reported • Survey duration: Not reported 	<p>Construct validity (discriminant validity)</p> <ul style="list-style-type: none"> • Comparison measure: DSM-5 diagnostic severity levels • Statistical analysis: Sensitivity, specificity, ROC curve • Criterion validity (concurrent validity) • Comparison measure: Compared to scores and measures of the frequency of marijuana use and marijuana-related consequences • Statistical analysis: Bivariate correlations between CUDIT-R composite scores and measures of the frequency of marijuana use per week and marijuana-related consequences 	<p>The CUDIT-R showed good internal consistency and concurrent validity with cannabis-related outcome measures including frequency of use, cannabis-related consequences, and total DSM-5 criteria endorsed. The CUDIT-R also showed evidence of discriminant validity across DSM-5 severity classifications, achieved high levels of sensitivity (0.929) and specificity (0.704), and excellent area under the receiver operating characteristics curve when using a cutoff score of six. All items displayed high levels of discrimination and varied in terms of difficulty and information provided</p>	5

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Schwartz, McNeely [46]	<ul style="list-style-type: none"> • Mean age (SD)=46.0 (14.7) • More participants were females ($n=1124$, 56.2%) compared to males ($n=874$, 43.7%) • More participants were African-American ($n=1112$, 55.6%), followed by White ($n=667$, 33.4%), Other race ($n=113$, 5.7%), multiracial ($n=66$, 3.3%), Asian ($n=35$, 1.8%) 	<p>Non-population-based study conducted in the inpatient clinical setting to examine the performance of a TAPS tool compared to the WHO ASSIST by conducting concurrent validity</p>	<p>TAPS</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n=2000$ • Response rate: Not reported • Survey duration: August 2014–April 2015 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: Comparison to the full WHO ASSIST as part of a large, multi-site study in eastern US primary care patients • Statistical analysis: Sensitivity, specificity, and PPV and NPV were calculated. ROC curves were computed, and the AUC was examined 	<p>For illicit drugs, sensitivities were ≥ 0.82 and specificities ≥ 0.92. The TAPS (at a cutoff of 1) had good sensitivity and specificity for moderate-risk tobacco use (0.83 and 0.97) and alcohol (0.83 and 0.74). Among illicit drugs, sensitivity was acceptable for moderate risk of marijuana (0.71), while it was low for all other illicit drugs and nonmedical use of prescription medications</p> <p>Specificities were 0.97 or higher for all illicit drugs and prescription medications</p>	5
Smith, Bennett [47]	<ul style="list-style-type: none"> • Age ranges from 18–25; mean age (SD)=21.0 (2.39), 35% ($N=3396$) were females • More participants were Whites ($n=5196$, 53%) followed by African American ($n=1563$, 16%), Latino ($n=1907$, 19%), and Other ($n=1147$, 12%) 	<p>Non-population-based study conducted in the outpatient clinical setting to examine the sensitivity and specificity of the SDSCrY for three past year criterion variables, AOD, AUD, or DUD, in predicting emerging adults (18–25) substance use disorders</p>	<p>GAIN Short SDSCrY</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n=804$ • Response rate: Not reported • Survey duration: Not reported 	<p>Criterion validity (predictive validity)</p> <ul style="list-style-type: none"> • Comparison measure: Compared to the past year AOD, AUD, and DUD • Statistical analysis: Sensitivity, specificity, and the ROC curve 	<p>Analyses revealed a high correlation between the SDSCrY screener and its longer parent scale ($r=0.95$, $p<0.001$). Sensitivity (83%) and specificity (95%) were highest at a cutoff score of two (AUC=94%) on the SDSCrY for any past year substance use disorder. Sensitivity (85%) was also high at a cutoff score of two on the SDSCrY for any past year alcohol disorder</p>	4

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Smith, Cheng [48]	<ul style="list-style-type: none"> • Age ranges from 21–86, median age=49, 54% were females • The majority of participants (63%) identified themselves as Black or African American 	<p>Non-population-based study conducted in the inpatient clinical setting to test the validity of a single question for unhealthy alcohol use to detect other drug use</p>	<p>SIP-DU</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 286$ • Response rate: 73% • Survey duration: October 2006–June 2007 	<p>Content validity</p> <ul style="list-style-type: none"> • Comparison measure: Not reported • Statistical analysis: Sensitivity, specificity, likelihood ratios, and AUC curve 	<p>The single-question screen at a cutoff of one or more times (the value considered a positive test for alcohol screening) was 67.6% sensitive (95% confidence interval (CI), 50.2–82.0%) and 64.7% specific (95% CI, 58.4–70.6%) for the detection of a current drug use disorder. It appeared slightly less sensitive (62.6%; 95% CI, 52.3–72.2%) and more specific (72.7%; 95% CI, 65.8–79.0%) for the detection of current drug use (although CIs overlapped). If oral fluid test results were considered, the sensitivity for detecting current drug use was lower (58.8%; 95% CI, 47.6–69.4%) and the specificity higher (80.3%; 95% CI, 72.5–86.7%)</p>	2

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Smith, Schmidt [49]	<p>Total study population</p> <ul style="list-style-type: none"> • Age: Age range from 21 to 86, mean (SD) = 49 (12.3) • Sex: <i>n</i> = 115 (54.2%) were women • Race: American Indian/Alaskan native: <i>n</i> = 8 (2.8%), Asian: <i>n</i> = 7 (2.4%), Black: <i>n</i> = 179 (62.6%), Native Hawaiian: <i>n</i> = 3 (1.1%), White: <i>n</i> = 49 (17.1%), unknown: <i>n</i> = 40 (14.0%), Hispanic: <i>n</i> = 46 (16.1%) • Education: Some high school: <i>n</i> = 81 (28.3%), high school graduate: <i>n</i> = 107 (37.4%), some college: <i>n</i> = 59 (20.6%), college graduate: <i>n</i> = 28 (9.8%), postgraduate education: <i>n</i> = 11 (3.9%) <p>Consent to oral fluid testing</p> <ul style="list-style-type: none"> • Age: Age ranges from 21 to 86, mean (SD) = 49.3 (12.8) • Sex: <i>n</i> = 135 (56.2%) were women • Race: American Indian/Alaskan native: <i>n</i> = 5 (2.1%), Asian: <i>n</i> = 5 (2.1%), Black: <i>n</i> = 153 (63.8%), Native Hawaiian: <i>n</i> = 2 (0.8%), White: <i>n</i> = 42 (17.4%), unknown: <i>n</i> = 33 (13.8%), Hispanic: <i>n</i> = 38 (15.8%) • Education: Some high school: <i>n</i> = 68 (28.4%), high school graduate: <i>n</i> = 86 (35.8%), some college: <i>n</i> = 50 (20.8%), college graduate: <i>n</i> = 26 (10.8%), postgraduate education: <i>n</i> = 10 (4.2%) 	<p>Non-population-based study conducted in the inpatient clinical setting to validate a single-question screening test for drug use and drug use disorders</p>	<p>SIP-DU</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: <i>n</i> = 286 • Response rate: 73% • Survey duration: October 2006–June 2007 	<p>Content validity</p> <ul style="list-style-type: none"> • Comparison measure: DAST-10 • Statistical analysis: Sensitivity, specificity, likelihood ratios, and AUC curve 	<p>The single screening question was 100% sensitive (95% confidence interval [CI], 90.6–100%) and 73.5% specific (95% CI, 67.7–78.6%) for the detection of a drug use disorder. It was less sensitive for the detection of self-reported current drug use (92.9%; 95% CI, 86.1–96.5%) and drug use detected by oral fluid testing or self-report (81.8%; 95% CI, 72.5–88.5%). Test characteristics were similar to those of the DAST-10 and were affected very little by participant demographic characteristics</p>	5

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Tarter and Kirisci [50]	<p>Group I: Alcohol and drug abusers</p> <ul style="list-style-type: none"> • Age: Mean (SD) = 40.9 (5.7) • Sex: Male: <i>n</i> = 56 (47.1%), female: <i>n</i> = 63 (52.9%) • Race: Euro-American: <i>n</i> = 104 (87.4%), African American: <i>n</i> = 14 (11.8%), Other: <i>n</i> = 1 (0.8%) <p>Group II: Normal controls</p> <ul style="list-style-type: none"> • Age: Mean (SD) = 41.18 (4.79) • Sex: Male: <i>n</i> = 59 (49.6%), female: <i>n</i> = 60 (50.4%) • Race: Euro-American: <i>n</i> = 108 (90.8%), African American: <i>n</i> = 11 (9.2%), Other: not reported 	<p>Non-population-based study conducted in the inpatient clinical setting to evaluate the structure and psychometric properties of the adult version of the DUSI, to determine the capacity of the DUSI to discriminate substance abusers from non-substance abusing individuals, and to test the sensitivity of the DUSI for detecting individual who qualify for a DSM-III-R diagnosis of abuse or dependence</p>	<p>DUSI</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: <i>n</i> = 238 (Group I); <i>n</i> = 299 (Group II) • Response rate: 73% • Survey duration: Not reported 	<p>Construct validity</p> <ul style="list-style-type: none"> • Comparison measure: Compared to other measures, such as Multidimensional Personality Questionnaire, Family Assessment Measure, and General Health Questionnaire • Statistical analysis: The scales of the DUSI were correlated with other measures to examine the construct validity of the scales 	<p>Each of the 10 DUSI domains is unidimensional. Inter-item, split half, and internal reliability ranged from good to excellent. A score of 4 or higher on the substance use domain correctly classified 80% of the substance abusers, whereas a score of 3 or less accurately detected 100% of the normal control subjects. These results demonstrate that the DUSI is a practical and psychometrically sound screening instrument</p>	5
Tiet, Leyva [51]	<ul style="list-style-type: none"> • Mean age (SD) = 62.2 (12.6) years • The majority of participants were males (95.25%) • Whites (53.7%) and other race (45.05%) 	<p>Non-population-based study conducted in the outpatient clinical setting to create and validate a two-item screen for drug use from the ASSIST (excluding tobacco and alcohol) and to improve the efficiency of screening of drug misuse in primary care</p>	<p>ASSIT drug</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: <i>n</i> = 1283 • Response rate: Not reported • Survey duration: Feb 2012–April 2014 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: MINI and the inventory of drug use consequences • Statistical analysis: Sensitivity, specificity 	<p>Based on the development sample, the ASSIST-Drug was 94.1% sensitivity and 98.6% specific for drug use disorders. Based on the validation sample, it was 95.4% sensitive and 87.8% specific</p>	4
Tiet, Leyva [52]	<ul style="list-style-type: none"> • Mean age (SD) = 62.2 (12.6) years • The majority of participants were males (95.25%) • Whites (53.7%) and Other race (45.05%) 	<p>Non-population-based study conducted in the outpatient clinical setting to examine the concurrent diagnostic accuracy of the SoDU in helping to detect marijuana use disorder</p>	<p>SoDU</p> <ul style="list-style-type: none"> • Specific substance use: Marijuana use • Sample size: <i>n</i> = 1283 • Response rate: Not reported • Survey duration: Feb 2012–April 2014 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: MINI • Statistical analysis: ROC analysis, sensitivity, specificity 	<p>The SoDU was 100% sensitive and 87.5% specific. When tested in subgroups of patients varying in age, gender, race/ethnicity, marital status, education level, and PTSD status, the SoDU maintained 100% sensitivity in all subgroups; specificity ranged from 76.26 to 94.34%</p>	4

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Tiet, Leyva [53]	<p>Total</p> <ul style="list-style-type: none"> • Mean age (SD) = 62.2 (12.6) years • The majority of participants were males (95.25%) • Whites (53.7%) and Other race (45.05%) • N = 969 (75.5%) completed higher than high school education <p>Validation</p> <ul style="list-style-type: none"> • Mean age (SD) = 62.63 (13.01) • The majority of participants were males (95.3%) • Whites (57.5%) • N = 485 (75.8%) completed higher than high school education 	<p>Non-population-based study conducted in the outpatient clinical setting to develop and validate the SoDU for diagnostic accuracy by conducting item performance analyses and to examine the performance</p>	<p>SoDU</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: Total: n = 1283, validation: n = 640 • Response rate: Not reported • Survey duration: Feb 2012–April 2014 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: The Mini-International Diagnostic Interview was used as the criterion for DUDs, and the Inventory of Drug Use Consequences was used as the criterion for NCDU • Statistical analysis: ROC analysis, sensitivity, specificity 	<p>The screening instrument was 100% sensitive and 93.73% specific for DUDs; when replicated in the second half of the sample, it was 92.31% sensitive and 92.87% specific. The screening instrument was 93.18% sensitive and 96.03% specific for NCDU; when replicated in the second half of the sample, it was 83.17% sensitive and 96.85% specific</p>	4
Tiet, Leyva [5]	<ul style="list-style-type: none"> • Mean age (SD) = 62.2 (12.6) years • The majority of participants were males (95.25%) • Whites (53.7%) and Other race (45.05%) 	<p>Non-population-based study conducted in the outpatient clinical setting to validate brief DAST and to test if the brief version of the DAST is practical for routine use in primary settings</p>	<p>DAST</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: n = 1283 • Response rate: Not reported • Survey duration: Feb 2012–April 2014 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: Two criterion measures were used: MINI for drug use disorders and the INDOC for negative consequences of drug use, which includes individuals who may or may not meet criteria for a drug use disorder • Statistical analysis: ROC analysis, sensitivity, specificity 	<p>The DAST-2 was 97% sensitive and 91% specific for DUDs in the development sample and 95% sensitive and 89% specific in the validation sample. It was highly sensitive and specific for DUD and negative consequences for drug use in subgroups of patients</p>	4
Tiet, Schutte [54]	<ul style="list-style-type: none"> • Mean age (SD) = 50.59 (9) years • The majority of participants were males (86.4%) • White (50%), African American/Black (31.8%), and Other race (9.9%) 	<p>Non-population-based study conducted in the outpatient clinical setting to conduct AUC, sensitivity, specificity, efficiency, PPV, and NPV of the PCL, PC-PTSD, and five abbreviated versions of the PCL in detecting PTSD for patients seeking treatment in substance use disorder specialty treatment</p>	<p>PTSD PCL-C, PTSD PCL-Bliese-4, PTSD-LS-2, PTSD PCL-LS-3, PTSD-PCL-LS-4, PTSD-PCL-LS-6, PC-PTSD</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Specific mental health: PTSD • Sample size: n = 242 • Response rate: Not reported • Survey duration: August 2003–December 2004 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: C-DIS-IV • Statistical analysis: AUC, cut-point sensitivity, specificity, PPV, NPV, efficient, test + % 	<p>Based on the C-DIS-IV, prevalence of PTSD was found to be 36.7 and 52.9% in the SUD and MH samples, respectively. The PCL, PC-PTSD, and five abbreviated versions of the PCL were found to have adequate psychometric properties for screening patients in SUD (AUC ranged from 0.80 to 0.86) and MH (AUC ranged from 0.77 to 0.80) outpatient treatment settings</p>	4

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Westermeyer, Crosby [55]	<ul style="list-style-type: none"> • Mean age (SD) = 30.3 (10.7) • Males (95.25%) and female (43.1%) 	<p>Non-population-based study conducted in the outpatient clinical setting to conduct current validity of the M-SAPS compared with three factors: psychiatric-behavioral problems, social-interpersonal problems, and addiction-dependence symptoms</p>	<p>M-SAPS</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 642$ • Response rate: Not reported • Survey duration: Not reported 	<p>Construct validity (convergent validity)</p> <ul style="list-style-type: none"> • Comparison measure: Not reported • Statistical analysis: Factor analysis <p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: Psychiatric rating, social problem rating scales, assistive use scales • Statistical analysis: Pearson correlation coefficients 	<p>All six psychiatric rating scales were most strongly correlated with psychiatric-behavioral problems. Additionally, all six comparisons with two social problem measures vs MMADST and Axis V coping scales across the three M-SAPS factors were highly significant. More severe psychosocial stressors in the last year and more substance dependence were associated with higher addictive use symptoms scores</p>	5
Wickersham, Azar [56]	<ul style="list-style-type: none"> • Mean age (SD) = 47.2 (8.3) • The majority of participants were males (70.1%) • White (19.6%), African American/Black (54.6%), and Other race (25.8%) 	<p>Non-population-based study and participants were recruited from a novel jail-release program to conduct the initial validation of the RODS among a sample of 97 newly incarcerated, HIV-positive individuals by comparing the MINI as the primary measure of opioid dependence</p>	<p>RODS</p> <ul style="list-style-type: none"> • Specific substance use: opioid use • Sample size: $n = 97$ • Response rate: Not reported • Survey duration: 2009–2011 	<p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: MINI • Statistical analysis: Concordance analysis, sensitivity, specificity, PPV, NPV 	<p>The RODS showed good-to-strong sensitivity (0.97), specificity (0.76), positive predictive value (0.69), and negative predictive value (0.98), while concordance analysis revealed moderate diagnostic agreement ($k = 0.67$)</p>	4
Woick, Stewart [57]*	<p>Mean age (SD) = 20 (3.1) years</p>	<p>Sub-population-based study and participants were undergraduate but did not specify the setting to conduct analysis of the internal structure of two versions of the SURPS and to conduct concurrent, discriminant, and incremental validity compared to other theoretically relevant personality and drug use criterion measures</p>	<p>SURPS</p> <ul style="list-style-type: none"> • Specific substance use: Opioid use • Sample size: $n = 195$ • Response rate: Not reported • Survey duration: Not reported 	<p>Construct validity (convergent and discriminant validity)</p> <ul style="list-style-type: none"> • Comparison measure: Personality scales of NEO-FFI, MAST, PH scores • Statistical analysis: Pearson correlation coefficients <p>Criterion validity (predictive validity)</p> <ul style="list-style-type: none"> • Comparison measure: Personality scales • Statistical analysis: Hierarchical regression analyses 	<p>The first set included the five subscales of the NEO-FFI and accounted for 5% ($R = .22$, $pb.08$, $\Delta R^2 = .05$) of the variance in MAST scores and 8% of the variance in PH scores ($R = 0.28$, $pb.01$, $\Delta R^2 = .08$)</p>	4

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
	<ul style="list-style-type: none"> • Mean age (SD) = 19.3 (3.1) years • More females (55.13%) than males (44.87%) 	<p>Non-population-based study and participants were recruited from Stony Brook University to conduct test-retest reliability and validity with respect to measuring personality vulnerability to reinforcement-specific substance use patterns</p>	<p>SURPS</p> <ul style="list-style-type: none"> • Specific substance use: Opioid use • Sample size: $n = 390$ • Response rate: Not reported • Survey duration: Not reported 	<p>Construct validity (convergent and discriminant validity)</p> <ul style="list-style-type: none"> • Comparison measure: DMQ • Statistical analysis: Pearson correlation coefficients 	<p>To test the equivalence of item measurements across samples, we applied a more stringent procedure in which all paths (i.e., factor loadings, variances, and covariances) in the model were constrained to be equal for both samples</p>	4
Zanis, McLellan [58]	<ul style="list-style-type: none"> • Mean age = 39 • Sex: All participants were males • 91% were African American, 8% Caucasian, and 1% Latino; • Education: Average education was 12.4 years 	<p>Non-population-based study conducted in the outpatient clinical setting with a sample of 98 homeless substance users awaiting temporary housing placement and shelter to examine the reliability and validity of the ASI</p>	<p>ASI</p> <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: $n = 98$ • Response rate: Not reported • Survey duration: Not reported 	<p>Construct validity (discriminant validity)</p> <ul style="list-style-type: none"> • Comparison measure: Compared the ASI composite score and severity ratings to other tests, including MAST, RAB alcohol, RAB drug, BDI, and SCL-90 • Statistical analysis: Correlations between ASI subscale and appropriate comparable test <p>Criterion validity (concurrent validity)</p> <ul style="list-style-type: none"> • Comparison measure: (1) Assessed by comparing a test measure to a known conceptually similar standard measure at the same point in time; in (2) a separate evaluation of the concurrent validity of the ASI drug scale, authors examined data from 25 subjects that had drug metabolites detected in a urine sample obtained during the first interview and compared this result with their self-reported use of drugs during the 30-day assessment period covered by the ASI interview • Statistical analysis: Correlations 	<p>Both composite score and severity rating measures were found to be quite independent with low intercorrelations. Three of the seven ASI composite scores were tested for and found to have moderate concurrent validity: alcohol ($r = 0.31$ to 0.36), drug ($r = 0.46$), and psychiatric ($r = 0.53$ to 0.66). Composite score interitem correlations were 0.70 or greater in each of the domains except for employment (0.50) and family (0.52)</p>	5

Table 1 (continued)

First author and publication year	Participant's characteristics	Study characteristics	Survey instrument/questionnaire characteristics	Validation methods	Key findings	ROB assessment
Zanis, McLellan [59]	<ul style="list-style-type: none"> • Mean age (SD) = 62 (6.95) years • The majority of participants were males (74.2%) • White (43.5%), African American/Black (56.5%) 	Non-population-based study conducted in the inpatient clinical setting to examine aspects of reliability, validity, and utility of ASI among individuals with severe and persistent mental illness and concurrent substance abuse disorders	ASI <ul style="list-style-type: none"> • Specific substance use: Not reported • Sample size: n = 62 • Response rate: Not reported • Survey duration: Not reported 	Criterion validity (concurrent validity) <ul style="list-style-type: none"> • Comparison measure: Compared with the urine sample at the conclusion of the first ASI interview • Statistical analysis: Not reported in detail 	The past 30-day self-reported drug use questions of the ASI have poor concurrent validity when compared with urine screens, which showed that the ASI has limited validity	5

Abbreviations (alphabetical): Alcohol or other drug use disorder (AOD), Addiction Severity Index (ASI), Alcohol, Smoking, and Substance Involvement Screening Test-Drug (ASSIST-Drug), area under the curve (AUC), alcohol use disorder (AUD), Alcohol Use Disorder and Associated Disabilities Interview Schedule-DSM-IV (AUDASIS-IV), Beck Depression Index (BDI), Cut down, Annoyed, Guilty, and Eye-Opener Substance Abuse Screening Tool (CAGE), the Computerized Diagnostic Interview Schedule for DSM-IV (C-DIS-IV), the Cannabis Use Disorders Identification Test-Revised (CUDIT-R), Dartmouth Assessment of Lifestyle Instrument (DALI), Drug Abuse Screening Test (DAST), Drinking Motives Questionnaire (DMQ), Diagnostic and Statistical Manual III-Revised (DSM-III-R), Diagnostic Statistical Manual-Guided Marijuana Inventory (DSM-G-MI), drug use disorder (DUD), Global Appraisal of Individual Needs (GAIN), Global Appraisal of Individual Needs Quick Version 3 (GAIN-Q3), Healthcare Effectiveness Data and Information Set (HEDIS), Inventory of Drug Use Consequences (INDUC), Lifetime Severity Index for Cocaine Use Disorder (LSJ-Cocaine), Michigan Alcoholism Screening Test (MAST), Mini International Neuropsychiatric Interview (MINI), Short-Form 12-Item Health Survey (SF-12), negative consequences of drug use (NCDU), neuroticism, extroversion, openness, agreeableness, and conscientiousness (NEO-FFI), National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), negative predictive value (NPV), Parents, Partners, Past, and Pregnancy Plus (4P's Plus), Personality Assessment Inventory Drug Problem Scale (PAIDRG), positive predictive value (PPV), PTSD Checklist 2 Item (PCL-LS-2), PTSD Checklist 3 Item (PCL-LS-3), PTSD Checklist 4 Item (PCL-LS-4), PTSD Checklist 6 Item (PCL-LS-6), PTSD Checklist-Civilian version (PCL-C), Primary Care-PTSD screen (PC-PTSD), Quantity Frequency Index (QFI), risk for AIDS behavior (RAA), Rapid Opioid Dependence Screen (RODS), Substance Abuse Subtle Screening Inventory-3 (SASSI-3), Structured Clinical Interview for DSM-III-R (SCID), Symptom Checklist-90 (SCL-90), screen of drug use (SoDU); self-rated substance abuse (SRSA); Single-Item Screening Questions (SISQs), Single Question Used from Short Inventory of Problems-Drug Use (SIP-DU), South Shore Problem Inventory-revised (SSPI), Substance Abuse Subtle Screening Inventory-2 (SASSI-2), Substance Dependence Severity Scale (SDSS), Substance Use Brief Screen (SUBS), Texas Christian University Drug Screen (TCUDS), Substance Use and Abuse Survey (SUAS), the Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS), the Chemical Use, Abuse, and Dependence (CUAD), the Drug Use Screening Inventory (DUSI), the Marijuana Screening Inventory (MSI-X), the Simple Screening Instrument (SSI), the Simple Screening Instrument for Substance Abuse (SSI-SA), the Longitudinal Substance Use Recall Instrument (LSUR), the Longitudinal Substance Use Recall Instrument for 12 Weeks instrument (LSUR-12), tobacco, alcohol, prescription medication, and other substance use (TAPS tool), receiver operating characteristics (ROC). *Studies conducted validity testing in two study populations/samples

Table 2 Participant recruitment strategies

Population based (<i>n</i> = 2)	Clinical (<i>n</i> = 35)	Non-population based & non-clinical (<i>n</i> = 9)
	<ul style="list-style-type: none"> • Inpatient (<i>n</i> = 19) • Outpatient (<i>n</i> = 11) • Both inpatient and outpatient (<i>n</i> = 5) 	<ul style="list-style-type: none"> • Veteran's administration shelter (<i>n</i> = 1) • Unspecified drug and alcohol program (<i>n</i> = 1) • Prison substance use treatment program (<i>n</i> = 1) • Transfer facility (<i>n</i> = 1) • Jail release program (<i>n</i> = 1) • NESARC-III subsample (<i>n</i> = 1) • University (<i>n</i> = 3)

of the measure's ability to screen for a specific SUD, including marijuana use [18, 21, 29, 40, 45], cocaine use [30], and opioid use [56]. Five studies validated measures for both substance use and mental health [23, 25, 40, 44, 54], of which one study used a measure for post-traumatic stress disorder (PTSD) screening [54]. The rest of the included studies did not specify a specific SUD for screening purposes but used a generic term for defining SUD. All measures and their frequency of use in the included studies are depicted in Fig. 2.

The majority of studies validated one single measure, of which five studies validated the Addiction Severity Index (ASI), [32, 35, 41, 58, 59] and one study validated drug use subscales of ASI [26, 37]. Five studies validated multiple survey measures:

- (1) Duncan et al. validated two survey measures: (a) CJDAT Co-Occurring Disorders Screening Instruments for any Mental Disorder (CODSI-MD) and (b) CJDAT Co-Occurring Disorders Screening Instruments for Severe Mental Disorder (CODSI-SMD) [25].
- (2) Ramsay et al. also validated two different survey measures: (a) The Lifetime Substance Use Recall Instrument (LSUR) and (b) the Longitudinal Substance Use Recall for 12 Weeks instrument (LSUR-12) [42].
- (3) Peters et al. and Tiet et al. also validated two measures: (a) The Substance Use Brief Screen (SUBS) and (b) the DAST [5, 41]. O'Hare et al. validated four different survey measures: (a) South Shore Problem Inventory-revised (SSPI), (b) self-rated substance abuse (SRSA), (c) quantity-frequency index for alcohol consumption (QFI), and (d) one-item index measuring the frequency of marijuana use [40].
- (4) Peters et al. validated five different survey measurements: (a) ASI-drug use subscales, (b) DAST, (c) Substance Abuse Subtle Screening Inventory-2

(SASSI-2), (d) SSI, and (e) Texas Christian University Drug Screen (TCUDS) [41].

- (5) Tiet et al. conducted validations of seven survey measures: (a) PTSD Checklist–Civilian version (PCL-C), (b) PTSD Checklist 4 Item (PCL-Bliese-4), (c) PTSD Checklist 2 Item (PCL-LS-2), (d) PTSD Checklist 3 Item (PCL-LS-3), (e) PTSD Checklist 4 Item (PCL-LS-4), (f) PTSD Checklist 6 Item (PCL-LS-6), and (g) Primary Care–PTSD screen (PC-PTSD) [54].

Two studies conducted survey measure validation in different study populations. One study conducted a preliminary exploration of the psychometric properties of the Substance Use Risk Profile Scale (SURPS) in 3 different populations: 195 undergraduate drinkers, 390 undergraduate students from Stony Brook University, and 4234 high school students in Canada [57]. In the second study, data were collected from two separate adult clinical samples — seriously mentally ill inpatients and patients presenting for evaluation at a chemical dependence program — to describe the rationale and test validity and reliability of the Chemical Use, Abuse, and Dependence Scale (CUAD) [36].

Comparison measures for validation

Several different types of measures were used as comparison for the purpose of validation. Higher quality comparison measures included items such as medical records, diagnoses, medical test results, or other SUD severity scales. A total of 10 studies conducted validity testing using at least one of these higher-quality comparison measures. Of these, three studies conducted criterion validity testing by comparing the following: (1) positive and negative 4P's Plus screens with positive and negative clinical assessment [22], (2) the Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS) with psychiatrist diagnosis [28], and (3) Dartmouth Assessment of Lifestyle Instrument (DALI) with clinician diagnosis [43]. The remaining two studies conducted validity testing by

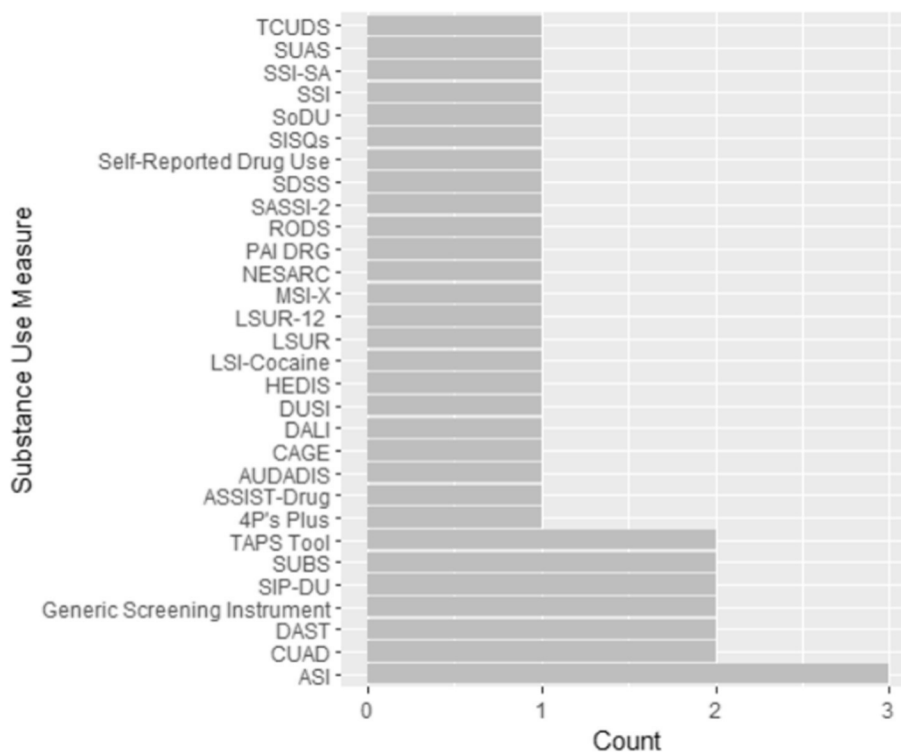


Fig. 2 Frequency of survey measures used in included studies. Abbreviations in order: Texas Christian University Drug Screen (TCUDS), Substance Use and Abuse Survey (SUAS), the Simple Screening Instrument for Substance Abuse (SSI-SA), the Simple Screening Instrument (SSI), screen of drug use (SoDU), single-item screening questions (SISQs), Substance Dependence Severity Scale (SDSS), Substance Abuse Subtle Screening Inventory-2 (SASSI-2), Rapid Opioid Dependence Screen (RODS), Personality Assessment Inventory Drug Problem Scale (PAI DRG), National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), the Marijuana Screening Inventory (MSI-X), the Longitudinal Substance Use Recall Instrument Recall for 12 Weeks instrument (LSUR-12), the Longitudinal Substance Use Recall Instrument (LSUR), Lifetime Severity Index for Cocaine Use Disorder (LSI-Cocaine), Healthcare Effectiveness Data and Information Set (HEDIS), the Drug Use Screening Inventory (DUSI), Dartmouth Assessment of Lifestyle Instrument (DALI), Cut down, Annoyed, Guilty, and Eye-Opener Substance Abuse Screening Tool (CAGE), the Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS), Alcohol, Smoking, and Substance Involvement Screening Test-Drug (ASSIST-Drug), Parents, Partners, Past, and Pregnancy Plus (4P's Plus), tobacco, alcohol, prescription medication, and other substance use (TAPS tool), Substance Use Brief Screen (SUBS), single question used from short inventory of problems-drug use (SIP-DU), Drug Abuse Screening Test (DAST), the Chemical Use, Abuse, and Dependence (CUAD), Addiction Severity Index (ASI)

comparing the following: (1) Substance Use and Abuse Survey (SUAS) with medical chart [34] and (2) CUDIT-R with DSM-T diagnostic severity levels [45]. Another study compared the CUAD-derived DSM-III-R substance use disorder diagnoses with the chart diagnosis determined by the unit psychiatrists for validation [36].

Furthermore, two studies validated their measures by comparing with diagnostic standards: (1) Compared Cut down, Annoyed, Guilty, and Eye-Opener Substance Abuse Screening Tool (CAGE) with SCID-generated drug use disorder diagnoses as the standard [24] and (2) Compared the Cannabis Use Disorders Identification Test Revised (CUDIT-R) with ICD-10 dependence diagnosis [39]. Three studies conducted validity testing by comparing with laboratory test results, including urine test [31, 58, 59] and saliva drug testing [38].

Four studies conducted validity testing by comparing other severity scales: (1) Criterion validity testing by comparing the Marijuana Screening Inventory (MSI-X) with three different severity rating scales and selected variables [18], (2) construct validity testing by comparing Personality Assessment Inventory Drug Problem Scale (PAI DRG) with ASI drug composite scores and severity ratings [33], (3) construct validity testing by comparing ASI with interviewer severity ratings and composite scores [35], and (4) concurrent validity of ASI drug scale and examined 25 participants who had drug metabolites detected in a urine sample obtained during the first interview and compare this result with their self-reported use of drugs during the 30-day assessment period in ASI interview [58].

Types of validity assessed and statistical analyses conducted

Two-thirds of the studies ($n=30$) included in this review examined criterion validity, specifically concurrent validity ($n=22$), predictive validity ($n=5$), and specification validity ($n=1$), and unspecified ($n=2$). Over half ($n=24$) studies conducted construct validity, specifically, convergent validity ($n=10$), discriminant validity ($n=6$), hypothesis testing validity ($n=1$), predictive validity ($n=2$), and factorial validity ($n=1$). Eight articles did not report specific types of construct validity. While three studies conducted content validity, none reported specific type of content validity [33, 37, 48]. Additionally, 11 studies conducted validity testing for multiple measures [18, 19, 29, 33, 36, 39, 40, 45, 55, 57, 58]. Ten studies investigated construct and criterion validity of a single survey measure [18, 19, 29, 36, 39, 45, 55, 57, 58], and one study conducted construct and content validity of a single survey measure [33].

Studies conducted the following statistical analyses for testing validity of survey measures: (1) sensitivity and specificity ($n=16$), (2) receiver operating characteristics (ROC) curve ($n=10$), (3) correlation coefficient ($n=9$), (4) Pearson correlation coefficient ($n=8$), and (5) positive predicted value (PPV) ($n=8$). Sensitivity and specificity were the most common statistical method for validation among studies examining construct validity and criterion validity.

Most studies showed strong evidence of validity or had strong significant associations with other measures for comparison. Studies that compared substance use measures with physician diagnoses or medical records showed strong overall validity. For example, Rosenberg et al. conducted ROC analysis for criterion validity and concluded that DALI functioned significantly better than traditional instruments for substance use disorders among psychiatric patients [43]. Compared with DAST-10, Short Inventory of Problems-Drug Use (SIP-DU) showed 100% sensitivity and 73.5% specificity for the detection for a drug use disorder. It was less sensitive at detecting self-reported current drug use (92.9%) and drug use detected by oral fluid testing or self-report (84.7%) [49]. However, studies demonstrated lack of validity for certain measures. For example:

- (1) Compared to urine screens, the ASI's questions about drug use in the past 30 days had poor concurrent validity, which suggested that the ASI has limited validity [59].
- (2) Correlations were not statistically significant among South Shore Problem Inventory-revised (SSPI) subscales and three other substance abuse indices, such as self-related substance abuse (SRSA), quantity-frequency index (QFI) for alcohol consumption,

and one-item index measuring the frequency of marijuana use [40].

- (3) Compared with oral fluid test results, using SIP-DU at a cut-off score (to be considered a positive test for alcohol screening) showed lower sensitivity and higher specificity for detecting current drug use [48].

Discussion

This systematic review found 46 studies conducted in the US between 1979 and 2021 that tested the validity of substance use/SUD measures. Two studies were population based [20, 28], while the rest were conducted in subpopulations or in clinical settings. Criterion validity and construct validity were the commonly used validation methods, and sensitivity and specificity were the most common statistical analyses for validation. More importantly, this review found that a myriad of survey measures was used to measure substance use/SUD. In addition, diverse methodologies were applied to measure validity, which makes comparability difficult. In general, most studies showed evidence of strong validity.

For example, among those articles included in this review, 46 studies tested the psychometric properties of 43 different substance use screening measures. Of them, 16 tested the validity of psychometric properties by comparing other self-reported survey measures, and one study conducted criterion validity by comparing different racial or ethnic groups of offenders [25]. Fourteen studies conducted concurrent validity by comparing measures with an external independent source or "gold standard," such as physician/clinician diagnosis, medical records or assessment, severity scales, or urine/saliva drug testing. Frequently, researchers rely on self-reported information on substance use to save time and cost and collect required information on a larger sample size than making comparison with a gold standard, such as a biological test or medical record.

The measures used in these studies varied greatly. The ASI, which was used most frequently in this review, was used in only five studies. Additionally, three articles specifically conducted validity testing for marijuana use. However, each of those studies used many diverse measures, such as MSI-X [18], the CUDIT-R [39, 45], screen of drug use (SoDU) [60], NESARC [60], a two-item brief screen with no instrument name reported [60], and one-item index measuring the frequency of marijuana use [40]. Multiple measures for one specific substance use might increase the likelihood of conflicting results, which can make it difficult to interpret and compare results across different studies. Thus, there is a need to adopt a standardized measure to ensure the results obtained are reliable and to be able to draw general conclusions.

In addition to the diverse measures, even the validation methods employed in the articles varied greatly. Although criterion and construct validity were the most commonly utilized validity measures, the specific type of criterion or construct validity varied among studies. For example, concurrent, predictive, and specification validity were reported as the three different types of criterion validity. Some studies employed multiple validation methods for a single survey measure, while others only used one. Moreover, different types of validity may achieve different objectives, which could explain the differences in statistical analyses of validation. This review also suggested that the statistical analyses used to test the validity of survey measures were diverse, with sensitivity and specificity being the most frequent analysis. Other statistical analyses such as ROC curve and correlation coefficient were also used to validate the survey measures.

Likewise, other differences were observed for demographic characteristics of participants. First, the validation of the substance use and SUD measures was primarily conducted in either inpatient or outpatient clinical settings, and only two studies were population based. Secondly, some studies had small sample sizes, which could significantly reduce the statistical power for finding differences between study groups. Moreover, some studies were occasionally limited to certain age or race/ethnicity groups, which could adversely affect the generalizability of findings. For example, several studies were restricted to White or Black/AA participants [6, 18, 21, 24, 25, 31, 32, 36, 39, 41, 58, 59]. In addition, information on race/ethnicity was missing from a few studies [19, 22, 27, 34, 55, 57]. Those studies might reflect racial disparities in SUD, as well as treatment for SUD. Although SUD is prevalent among all racial groups, the burden of disease is disproportionate among Black people, and treatment of SUD is less available for Black people [61]. Three studies were limited to either males or females only [22, 41, 58]. These studies provide valuable validation in the respective populations and may prove useful in other populations. However, further validation is needed in diverse populations for these measures to be generalizable.

SUD often co-occurs with many other physical and mental health conditions. Previous studies have shown a high co-occurrence and the increased risk of mental health disorders among individuals with SUD, which can be observed in clinical samples [62, 63]. In this review, only five studies validated measures for both substance use and mental health disorders. Results from studies assessing substance use and mental health simultaneously can help inform integrated treatment interventions by connecting individuals with additional service providers who can provide specialized

services to treat the physical and emotional elements of mental health and SUD [64]. Additional advantages of assessing co-occurring substance use and mental health include decreased hospitalization, fewer arrests, and increased housing stability [64]. More importantly, assessing co-occurring substance use and mental health disorders in population research can identify the barriers and disparities of treatment access, including race/ethnicity [65] and low treatment utilization among individuals with only substance use or only mental health disorders [66, 67].

Although this review adhered to the PRISMA guidelines, it is not without limitations. It was limited to studies conducted in the US, and studies in other countries were not included. Research shows that significant contextual differences, such as burden of substance use disorders, cultural norms, legal frameworks, healthcare systems, and societal attitudes towards substance use, can vary widely across countries, potentially influencing the reliability and applicability of measures developed and validated in one context when applied to another [1–3]. Our focus on US-based studies aims to ensure that the measures reviewed are relevant and applicable to the US population, providing a more accurate and context-specific assessment of substance use and SUDs.

Although a rigorous search strategy was implemented, our search was limited to library databases. As such, key clinical surveys were used in hospitals or other specialty clinical settings that were not published in peer-reviewed journals and may be missing from our review. Additionally, our objectives were to summarize the validity of measures to assess the prevalence of substance use and SUD in the US estimated in population and sub-population-based surveys. Therefore, we did not specifically review the best clinical practices for survey administration in the clinical setting. Findings highlight the need to evaluate substance use surveys in a population-based setting to identify a valid survey for use across population-based surveys. The consistent use of one survey may provide for more accurate comparisons across populations. However, the main limitation of this review is that the articles included in this review are missing information about demographic characteristics, such as the distribution of race and ethnicity groups in the study population, and only 5 studies in this review reported education level of the participants [33, 37–39, 49, 58]. The variation in the accuracy of self-reported data about substance use depends on education and socioeconomic status [68]. The majority of studies included in this review did not report the response rate or the survey duration. Lastly, our analyses relied only on peer-review studies, and our review did not include internal

studies that may have been conducted in large surveys, such as NSDUH.

This study has several strengths. To our knowledge, it is the first systematic review to summarize the validity of substance use/SUD measures used in questionnaires or instruments among US adults. This review has included 43 years of data among nine different literature databases. In addition, it has also included “gray literature” such as theses and Google Scholar, which can make significant contributions to systematic reviews by minimizing publication bias, enabling a more impartial assessment of the evidence, and publicizing null or negative findings [69]. Another strength of the study is that the methodologic quality of validation studies was assessed by an adapted risk-of-bias tool, created especially for this assessment. Lastly, while previous reviews have explored the instruments used to assess substance use and the identification of disorders [7, 8], this review uniquely concentrates on a comprehensive evaluation of the psychometric properties of measures assessing a broader spectrum of substances. This review aimed to distinguish from previous research, highlighting the diversity and specificity of instruments in current use, their applicability in various population and sub-population surveys, and the critical need for standardized, short, and versatile measures.

The findings of this review have several key implications. The study demonstrates that survey questions can be used to assess the prevalence of SUD in specific populations. However, most studies used different measures suggesting there was no consensus on the best measure to use for assessing the prevalence of substance use and SUD. This lack of common measures illustrates the difficulty in assessing SUD in short surveys, especially for specific substances. Similar to a global measure of psychological distress that is used to indicate nonspecific psychological distress [70], a measure is needed for measuring SUD in population-based studies. Only 5 out of 46 studies were conducted in population or sub-population-based settings. Therefore, more research needs to be conducted to validate these measures in population-based settings to confirm their sensitivity and specificity. Additionally, more studies need to validate measures using a “gold standard,” such as an outside reliable measure, because comparing with self-reported substance use can result in misclassification bias. Therefore, this systematic review illustrates a critical need to develop short measures for assessing SUD that do not require lengthy, time-consuming data collection that would be difficult to incorporate into population-based surveys assessing a multitude of health dimensions.

Conclusion

This systematic review summarized the validity of measures used to assess the prevalence of substance use and SUD in the US estimated in general population surveys and other population-based settings. Among the 46 studies included, this review demonstrated that a myriad of survey measures were used to assess substance use and SUD, and diverse methodologies were used to measure validity. This information suggests a lack of standardized, comparative survey measures in assessing the prevalence of substance use and SUD among US adults. This inconsistency makes it difficult to recommend the best measures to use in US surveys and highlights the need to develop better summary measures. Very few studies in this review were conducted in general population settings, which suggests that more research is needed to validate substance use measures in such settings. Although SUD is prevalent among all racial/ethnicity, age, and gender/sex groups in the US, and studies in this review provided valuable validation in the respective populations, further validation is needed in diverse populations. Thus, future validation research needs to be conducted in population-based settings to adopt a standardized measure for substance use and SUD that can inform interventions aimed to detect and manage problems associated with substance use and SUD and prevent avoidable premature US deaths.

Abbreviations

SUD	Substance use disorder
US	United States
DSM	<i>Diagnostic and Statistical Manual of Mental Disorders</i>
NSDUH	National Survey of Drug Use and Health
NAVIPPRO™	National Addictions Vigilance Intervention and Prevention Program
ASI-MV®	Addiction Severity Index-Multimedia Version®
DAST	Drug Abuse Screening Test
ASSIST	Alcohol, Smoking, and Substance Involvement Screening Test
TAPS	Tobacco, alcohol, prescription medication, and other substance use
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
BRFSS	Behavioral Risk Factor Surveillance System
AA	African American
SSI-SA	Simple Screening Instrument for Substance Abuse
NESARC-III	National Epidemiologic Survey on Alcohol and Related Conditions-III
PRISM-5	Psychiatric Research Interview for Substance and Mental Disorders, DSM-5 version
PTSD	Post-traumatic stress disorder
ASI	Addiction Severity Index
CODSI-MD CJDAT	Co-Occurring Disorders Screening Instruments for any Mental Disorder
CODSI-SMD CJDAT	Co-Occurring Disorders Screening Instruments for Severe Mental Disorder
LSUR	The Lifetime Substance Use Recall instrument
LSUR-12	The Longitudinal Substance Use Recall for 12 Weeks instrument
SSPI	South Shore Problem Inventory-revised

SRSA	Self-rated substance abuse
QFI	Quantity-frequency index for alcohol consumption
SASSI-2	Substance Abuse Subtle Screening Inventory-2
TCUDS	Texas Christian University Drug Screen
PCL-C	PTSD Checklist–Civilian version
PCL-Bliese-4	PTSD Checklist 4 Item
PCL-LS-2	PTSD Checklist 2 Item
PCL-LS-3	PTSD Checklist 3 Item
PCL-LS-4	PTSD Checklist 4 Item
PCL-LS-6	PTSD Checklist 6 Item
PC-PTSD	Primary Care-PTSD screen
SURPS	Substance Use Risk Profile Scale
CUAD	Chemical Use, Abuse, and Dependence Scale
AUDADIS	Associated Disabilities Interview Schedule
DALI	Dartmouth Assessment of Lifestyle Instrument
SUAS	Substance Use and Abuse Survey
CAGE	Cut down, Annoyed, Guilty, and Eye-Opener Substance Abuse Screening Tool
CUDIT-R	Cannabis Use Disorders Identification Test-Revised
MSI-X	Marijuana Screening Inventory
PAI DRG	Personality Assessment Inventory Drug Problem Scale
ROC	Receiver operating characteristics
PPV	Positive predicted value
SIPDU	Short Inventory of Problems-Drug Use
SRSA	Self-related substance abuse
QFI	Quantity-frequency index
SoDU	Screen of drug use

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13643-024-02536-x>.

Additional file 1: Supplementary Table 1. Search term list for each database. Figure 2. Bar Graph of Survey Measures Validated by Included Studies

Authors' contributions

YT and RB are responsible for research conception, design, and coordination of entire manuscript; YT is responsible for collection of data, literature database search, and article retrieval, for writing the initial manuscript, and for revising the manuscript; YT, EC, RM, NW, and EO are responsible for the literature review and assessment and interpretation of results and for reviewing and revising manuscripts; and GS, SDH, and RB are responsible for interpretation of results and for reviewing and revising manuscripts. All authors provided critical feedback to the manuscript and approved the final manuscript draft for submission.

Financial disclosure statement

No authors have financial relationships relevant to this article to disclose.

Funding

The review is being funded by the West Virginia Department of Health and Human Resources (WVDHHR).

Availability of data and materials

Table 1 contains the extracted data, and supplementary file contains the search strategy.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Epidemiology and Biostatistics, School of Public Health, West Virginia University, Morgantown, WV 26501, USA. ²Department of Pharmaceutical Systems and Policy, School of Pharmacy, West Virginia University, Morgantown, WV, USA. ³Health Affairs Institute, West Virginia University, Morgantown, WV, USA. ⁴Present Address: Highway Safety Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁵Present Address: Biological Sciences Division, University of Chicago, Chicago, USA, IL. ⁶Hartley and Associates, WV, Morgantown, USA.

Received: 12 October 2023 Accepted: 17 April 2024

Published online: 27 June 2024

References

- Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2021 National Survey on Drug Use and Health. Rockville, MD; 2022. Report No.: HHS Publication No. PEP22-07-01-005.
- Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2021 National Survey on Drug Use and Health. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2022.
- Butler SF, Budman SH, Licari A, Cassidy TA, Lioy K, Dickinson J, et al. National Addictions Vigilance Intervention and Prevention Program (NAVIPPRO): a real-time, product-specific, public health surveillance system for monitoring prescription drug abuse. *Pharmacoepidemiol Drug Saf.* 2008;17(12):1142–54.
- Kacha-Ochana A, Jones C, Green J, Dunphy C, Dailey T, Robbins R, et al. Characteristics of adults aged ≥ 18 years evaluated for substance use and treatment planning - United States, 2019. *MMWR Morb Mortal Wkly Rep.* 2022;71:749–56.
- Tiet QQ, Leyva YE, Moos RH, Smith B. Diagnostic accuracy of a two-item Drug Abuse Screening Test (DAST-2). *Addict Behav.* 2017;74:112–7.
- Carter G, Yu Z, Aryana Bryan M, Brown JL, Winhusen T, Cochran G. Validation of the tobacco, alcohol, prescription medication, and other substance use (TAPS) tool with the WHO alcohol, smoking, and substance Involvement screening test (ASSIST). *Addictive Behaviors.* 2022;126:107178. <https://doi.org/10.1016/j.addbeh.2021.107178>.
- Stewart RE, Cardamone NC, Schachter A, Becker C, McKay JR, Becker-Haimes EM. A systematic review of brief, freely accessible, and valid self-report measures for substance use disorders and treatment. *Drug Alcohol Depend.* 2023;243:109729.
- Boness CL, Carlos Gonzalez J, Sleep C, Venner KL, Witkiewitz K. Evidence-based assessment of substance use disorder. *Assessment.* 2023;31(1):168–90.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, The PRISMA, et al. statement: an updated guideline for reporting systematic reviews. *BMJ.* 2020;2021:372.
- Doyle SR, Donovan DM. A validation study of the alcohol dependence scale. *J Stud Alcohol Drugs.* 2009;70(5):689–99.
- Tevik K, Bergh S, Selbæk G, Johannessen A, Helvik A-S. A systematic review of self-report measures used in epidemiological studies to assess alcohol consumption among older adults. *PLoS ONE.* 2021;16(12):e0261292.
- Singh PN, Khieng S, Yel D, Nguyen D, Job JS. Validity and reliability of survey items and pictograms for use in a National Household Survey of Tobacco Use in Cambodia. *Asia Pacific Journal of Public Health.* 2013;25(5):455–535.
- Szklo AS, Iglesias RM, Stoklosa M, Figueiredo VC, Welding K, de Souza Junior PRB, et al. Cross-validation of four different survey methods used to estimate illicit cigarette consumption in Brazil. *Tob Control.* 2022;31(1):73.
- Pierannunzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. *BMC Med Res Methodol.* 2013;13(1):49.
- StataCorp. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC; 2021.
- Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability

- coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol.* 2013;13:61. <https://doi.org/10.1186/1471-2288-13-61>.
17. Centers for Disease Control and Prevention. Methods, validity, and reliability bibliography. Selected articles related to BRFSS and other self-reported data. 2023 [Available from: <https://www.cdc.gov/brfss/publications/mvr.html>].
 18. Alexander D, Leung P. The Marijuana Screening Inventory (MSI-X): concurrent, convergent and discriminant validity with multiple measures. *Am J Drug Alcohol Abuse.* 2006;32(3):351–78.
 19. Appleby L, Dyson V, Altman E, McGovern MP, Luchins DJ. Utility of the Chemical Use, Abuse, and Dependence Scale in screening patients with severe mental illness. *Psychiatr Serv.* 1996;47(6):647–9.
 20. Boothroyd RA, Peters RH, Armstrong MI, Rynearson-Moody S, Caudy M. The psychometric properties of the Simple Screening Instrument for Substance Abuse. *Eval Health Prof.* 2015;38(4):538–62.
 21. Broderick KB, Richmond MK, Fagan J, Long AW. Pilot Validation of a Brief Screen Tool for Substance Use Detection in Emergency Care. *J Emerg Med.* 2015;49(3):369–74.
 22. Chasnoff IJ, Wells AM, McGourty RF, Bailey LK. Validation of the 4P's Plus screen for substance use in pregnancy validation of the 4P's Plus. *J Perinatol.* 2007;27(12):744–8.
 23. Dennis ML, Davis JP. Screening for more with less: validation of the Global Appraisal of Individual Needs Quick v3 (GAIN-Q3) screeners. *J Subst Abuse Treat.* 2021;126:108414.
 24. Dezman ZDW, Gorelick DA, Soderstrom CA. Test characteristics of a drug CAGE questionnaire for the detection of non-alcohol substance use disorders in trauma inpatients. *Injury.* 2018;49(8):1538–45.
 25. Duncan A, Sacks S, Melnick G, Cleland CM, Pearson FS, Coen C. Performance of the CJDATS Co-Occurring Disorders Screening Instruments (CODSIs) among minority offenders. *Behav Sci Law.* 2008;26(4):351–68.
 26. Han BH, Sherman SE, Link AR, Wang B, McNeely J. Comparison of the Substance Use Brief Screen (SUBS) to the AUDIT-C and ASSIST for detecting unhealthy alcohol and drug use in a population of hospitalized smokers. *J Subst Abuse Treat.* 2017;79:67–74.
 27. Harris AH, Ellerbe L, Phelps TE, Finney JW, Bowe T, Gupta S, et al. Examining the specification validity of the HEDIS Quality Measures for Substance Use Disorders. *J Subst Abuse Treat.* 2015;53:16–21.
 28. Hasin DS, Greenstein E, Aivadyan C, Stohl M, Aharonovich E, Saha T, et al. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5): procedural validity of substance use disorders modules through clinical re-appraisal in a general population sample. *Drug Alcohol Depend.* 2015;148:40–6.
 29. Hasin DS, Keyes KM, Alderson D, Wang S, Aharonovich E, Grant BF. Cannabis withdrawal in the United States: results from NESARC. *J Clin Psychiatry.* 2008;69(9):1354–63.
 30. Hser Y, Shen H, Grella C, Anglin MD. Lifetime severity index for cocaine use disorder (LSI-Cocaine): a predictor of treatment outcomes. *J Nerv Ment Dis.* 1999;187(12):742–50.
 31. Jackson CT, Covell NH, Frisman LK, Essock SM. Validity of self-reported drug use among people with co-occurring mental health and substance use disorders. *J Dual Diagn.* 2005;1(1):49–63.
 32. Joyner LM, Wright JD, Devine JA. Reliability and validity of the Addiction Severity Index among homeless substance misusers. *Subst Use Misuse.* 1996;31(6):729–51.
 33. Kellogg SH, Ho A, Bell K, Schluger RP, McHugh PF, McClary KA, et al. The Personality Assessment Inventory Drug Problems Scale: a validity analysis. *J Pers Assess.* 2002;79(1):73–84.
 34. Kupetz K, Klagsbrun M, Wisoff D, La Rosa J, Davis DI. The acceptance and validity of the Substance Use and Abuse Survey (SUAS). *J Drug Educ.* 1979;9(2):163–80.
 35. Leonhard C, Mulvey K, Gastfriend DR, Shwartz M. The Addiction Severity Index: a field study of internal consistency and validity. *J Subst Abuse Treat.* 2000;18(2):129–35.
 36. McGovern MP, Morrison DH. The Chemical Use, Abuse, and Dependence Scale (CUAD): rationale, reliability, and validity. *J Subst Abuse Treat.* 1992;9(1):27–38.
 37. McNeely J, Cleland CM, Strauss SM, Palamar JJ, Rotrosen J, Saitz R. Validation of self-administered single-item screening questions (SISQs) for unhealthy alcohol and drug use in primary care patients. *J Gen Intern Med.* 2015;30(12):1757–64.
 38. McNeely J, Strauss SM, Saitz R, Cleland CM, Palamar JJ, Rotrosen J, et al. A brief patient self-administered substance use screening tool for primary care: two-site validation study of the Substance Use Brief Screen (SUBS). *Am J Med.* 2015;128(7):784.e9–19.
 39. Miele GM, Carpenter KM, Cockerham MS, Trautman KD, Blaine J, Hasin DS. Substance Dependence Severity Scale: reliability and validity for ICD-10 substance use disorders. *Addict Behav.* 2001;26(4):603–12.
 40. O'Hare T, Cutler J, Sherrer MV, McCall TM, Dominique KN, Garlick K. Co-occurring psychosocial distress and substance abuse in community clients: initial validity and reliability of self-report measures. *Community Ment Health J.* 2001;37(6):481–7.
 41. Peters RH, Greenbaum PE, Steinberg ML, Carter CR, Ortiz MM, Fry BC, et al. Effectiveness of screening instruments in detecting substance use disorders among prisoners. *J Subst Abuse Treat.* 2000;18(4):349–58.
 42. Ramsay CE, Abedi GR, Marson JD, Compton MT. Overview and initial validation of two detailed, multidimensional, retrospective measures of substance use: the Lifetime Substance Use Recall (LSUR) and Longitudinal Substance Use Recall for 12 Weeks (LSUR-12) instruments. *J Psychiatr Res.* 2011;45(1):83–91.
 43. Rosenberg SD, Drake RE, Wolford GL, Mueser KT, Oxman TE, Vidaver RM, et al. Dartmouth Assessment of Lifestyle Instrument (DALI): a substance use disorder screen for people with severe mental illness. *Am J Psychiatry.* 1998;155(2):232–8.
 44. Salyers MP, Bosworth HB, Swanson JW, Lamb-Pagone J, Osher FC, Salyers MP, et al. Reliability and validity of the SF-12 health survey among people with severe mental illness. *Med Care.* 2000;38(11):1141–50.
 45. Schultz NR, Bassett DT, Messina BG, Correia CJ. Evaluation of the psychometric properties of the cannabis use disorders identification test - revised among college students. *Addict Behav.* 2019;95:11–5.
 46. Schwartz RP, McNeely J, Wu LT, Sharma G, Wahle A, Cushing C, et al. Identifying substance misuse in primary care: TAPS tool compared to the WHO ASSIST. *J Subst Abuse Treat.* 2017;76:69–76.
 47. Smith DC, Bennett KM, Dennis ML, Funk RR. Sensitivity and specificity of the gain short-screener for predicting substance use disorders in a large national sample of emerging adults. *Addict Behav.* 2017;68:14–7.
 48. Smith PC, Cheng DM, Allensworth-Davies D, Winter MR, Saitz R. Use of a single alcohol screening question to identify other drug use. *Drug Alcohol Depend.* 2014;139:178–80.
 49. Smith PC, Schmidt SM, Allensworth-Davies D, Saitz R. A single-question screening test for drug use in primary care. *Arch Intern Med.* 2010;170(13):1155–60.
 50. Tarter RE, Kirisci L. The Drug Use Screening Inventory for adults: psychometric structure and discriminative sensitivity. *Am J Drug Alcohol Abuse.* 1997;23(2):207–19.
 51. Tiet QQ, Leyva Y, Moos RH, Smith B. Diagnostic accuracy of a two-item screen for drug use developed from the alcohol, smoking and substance involvement screening test (ASSIST). *Drug Alcohol Depend.* 2016;164:22–7.
 52. Tiet QQ, Leyva YE, Browne K, Moos RH. Screen of drug use: diagnostic accuracy for cannabis use disorder. *Addict Behav.* 2019;95:184–8.
 53. Tiet QQ, Leyva YE, Moos RH, Frayne SM, Osterberg L, Smith B. Screen of drug use: diagnostic accuracy of a New Brief Tool for Primary Care. *JAMA Intern Med.* 2015;175(8):1371–7.
 54. Tiet QQ, Schutte KK, Leyva YE. Diagnostic accuracy of brief PTSD screening instruments in military veterans. *J Subst Abuse Treat.* 2013;45(1):134–42.
 55. Westermeyer J, Crosby R, Nugent S. The Minnesota Substance Abuse Problems Scale Psychometric analysis and validation in a clinical population. *Am J Addict.* 1998;7(1):24–34.
 56. Wickersham JA, Azar MM, Cannon CM, Altice FL, Springer SA. Validation of a brief measure of opioid dependence: the Rapid Opioid Dependence Screen (RODS). *J Correct Health Care.* 2015;21(1):12–26.
 57. Woicik PA, Stewart SH, Pihl RO, Conrod PJ. The Substance Use Risk Profile Scale: a scale measuring traits linked to reinforcement-specific substance use profiles. *Addict Behav.* 2009;34(12):1042–55.
 58. Zanis DA, McLellan AT, Cnaan RA, Randall M. Reliability and validity of the Addiction Severity Index with a homeless sample. *J Subst Abuse Treat.* 1994;11(6):541–8.
 59. Zanis DA, McLellan AT, Corse S. Is the Addiction Severity Index a reliable and valid assessment instrument among clients with severe and

- persistent mental illness and substance abuse disorders? *Community Ment Health J.* 1997;33(3):213–27.
60. Broderick KB, Kaplan B, Martini D, Caruso E. Emergency physician utilization of alcohol/substance screening, brief advice and discharge: a 10-year comparison. *J Emerg Med.* 2015;49(4):400–7.
 61. Farahmand P, Arshed A, Bradley MV. Systemic racism and substance use disorders. *Psychiatr Ann.* 2020;50:494–8.
 62. Lieb M, Wittchen H-U, Palm U, Apelt SM, Siegert J, Soyka M. Psychiatric comorbidity in substitution treatment of opioid-dependent patients in primary care: prevalence and impact on clinical features. *Heroin Addiction and related clinical problems.* 2010;12(4):5–16.
 63. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry.* 2005;62(6):617–27.
 64. Substance Abuse and Mental Health Services Administration. The case for screening and treatment of co-occurring disorders 2022 [Available from: <https://www.samhsa.gov/co-occurring-disorders>].
 65. Hatzenbuehler ML, Keyes KM, Narrow WE, Grant BF, Hasin DS. Racial/ethnic disparities in service utilization for individuals with co-occurring mental health and substance use disorders in the general population: results from the national epidemiologic survey on alcohol and related conditions. *J Clin Psychiatry.* 2008;69(7):1112–21.
 66. Curran GM, Sullivan G, Williams K, Han X, Collins K, Keys J, et al. Emergency department use of persons with comorbid psychiatric and substance abuse disorders. *Ann Emerg Med.* 2003;41(5):659–67.
 67. Verduin ML, Carter RE, Brady KT, Myrick H, Timmerman MA. Health service use among persons with comorbid bipolar and substance use disorders. *Psychiatr Serv.* 2005;56(4):475–80.
 68. Hunt DE, Kling R, Almozlino Y, Jalbert S, Chapman MT, Rhodes W. Telling the Truth About Drug Use: How Much Does It Matter? *Journal of Drug Issues.* 2015;45(3):314–29.
 69. Benzies KM, Premji S, Hayden KA, Serrett K. State-of-the-evidence reviews: advantages and challenges of including grey literature. *Worldviews Evid Based Nurs.* 2006;3(2):55–61.
 70. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med.* 2002;32(6):959–95.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.