

Accuracy of Three Screening Tools for Prenatal Substance Use

Victoria H. Coleman-Cowger, PhD, Emmanuel A. Oga, MD, MPH, Erica N. Peters, PhD, Kathleen E. Trocin, MPH, Bartosz Koszowski, PharmD, PhD, and Katrina Mark, MD

OBJECTIVE: To compare and evaluate the accuracy of three screening tools in identifying illicit drug use and prescription drug misuse among a diverse sample of pregnant women.

METHODS: This prospective cross-sectional study enrolled a consecutive sample of 500 pregnant women, stratified by trimester, receiving care in two prenatal clinical settings in Baltimore, Maryland, from January 2017 to January 2018. All participants were administered three index tests: 4P's Plus, NIDA Quick Screen-ASSIST (Modified Alcohol, Smoking and Substance Involvement Screening Test), and the SURP-P (Substance Use Risk Profile-Pregnancy) scale, and administered reference tests (urine and hair drug testing) at the in-person baseline visit. To assess test-retest reliability of the index tests, screening tool administrations were repeated 1 week later by telephone. For each screening tool, sensitivity, specificity, positive predictive value, negative predictive value and test-retest reliability were computed. Results were stratified by age, race, and trimester of pregnancy.

RESULTS: Of the 500 enrolled pregnant women, 494 completed the index screening tools, 497 completed

reference testing, and 453 underwent test-retest analysis. For the 4P's Plus, sensitivity=90.2% (84.5, 93.8), and specificity=29.6% (24.4, 35.2). For the NIDA Quick Screen-ASSIST, sensitivity=79.7% (71.2, 84.2), and specificity=82.8% (78.1, 87.1). For the SURP-P, sensitivity=92.4% (87.6, 95.8) and specificity=21.8% (17.4, 27.2). Test-retest reliability (phi correlation coefficients) was 0.84, 0.77, and 0.79 for the 4P's Plus, NIDA Quick Screen-ASSIST and the SURP-P, respectively. For all screening tools, there were differences in validity indices by age and race, but no differences by trimester.

CONCLUSION: The SURP-P and 4P's Plus had high sensitivity and negative predictive values, making them more ideal screening tests than the NIDA Quick Screen-ASSIST. A clear recommendation for a clinically useful screening tool for prenatal substance use is crucial to allow for prompt and appropriate follow-up and intervention.

(*Obstet Gynecol* 2019;133:952-61)

DOI: 10.1097/AOG.0000000000003230

Substance use during pregnancy is a significant public health issue in the United States, with increasing illicit drug use observed among pregnant women from 2015 to 2017.¹ According to the 2016 National Survey on Drug Use and Health, self-reported past-month illicit drug use (inclusive of nonmedical use of prescription drugs) is 14.3% among pregnant adolescents ages 15–17 years, 10.1% among pregnant young adults (18–25 years), and 5.6% among pregnant adults (26–44 years).² These rates vary by trimester, with substance use typically decreasing over the course of pregnancy.³ Substance use during pregnancy may lead to multiple health and social problems for both mother and child, including miscarriage, stillbirth, low birth weight, prematurity, physical malformations, and neurologic damage.⁴

The American College of Obstetricians and Gynecologists strongly recommends substance use screening for pregnant women,⁵ and a 2012 expert panel convened by the Centers for Disease Control

From the Battelle Memorial Institute and the University of Maryland Medical School, Baltimore, Maryland.

The research reported in this article was supported by the National Institute on Drug Abuse of the National Institutes of Health under Award Number R01DA041328 (PI-Coleman-Cowger). The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

Each author has confirmed compliance with the journal's requirements for authorship.

Corresponding author: Victoria H. Coleman-Cowger, PhD, The Emmes Corporation, 401 N. Washington Street, 7th Floor, Rockville, MD 20850; email: vcolemancowger@emmes.com.

Financial Disclosure

The authors did not report any potential conflicts of interest.

© 2019 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ISSN: 0029-7844/19

and Prevention concluded that prenatal substance use screening should be universal.⁶ Although many health care providers use biological testing to determine use, a positive urine toxicology does not provide any context regarding temporality of use or indications of problematic use. Although validated alcohol and tobacco screening tools have been recommended by the U.S. Preventive Services Task Force, no specific substance use screening tool has been recommended for use with pregnant women to identify prescription drug misuse or other illicit drug use.

In finding a substance use screening tool that is efficacious for pregnant women, it is particularly important to ensure that it works in all subgroups, as studies show varying substance use by age,² race,⁷ and trimester.³ The primary aim of this study is to compare and validate screening tools within prenatal clinics to determine validity in identifying illicit drug use and prescription drug misuse among a diverse sample of pregnant women. The three screening tools used in this study—4P's Plus,⁸ NIDA Quick Screen-ASSIST (Modified Alcohol, Smoking and Substance Involvement Screening Test),⁹ and SURP-P (Substance Use Risk Profile-Pregnancy) scale¹⁰—were chosen because they are brief and are the only ones listed by the World Health Organization (WHO) to have been validated (though not all with a pregnant population) and to allow for screening of multiple substances.⁴

Our end goal is to provide evidence-based guidance to clinicians and encourage adoption of the recommended screening tool(s) into clinical practice. This is a first step in offering the U.S. Preventive Services Task Force “good quality research”¹¹ that screens pregnant women for prescription and illicit drug use and, thus, provides evidence for a recommendation for a standardized substance use screening tool.

METHODS

In this cross-sectional prospective study, we enrolled pregnant women presenting to two prenatal clinics in Baltimore, Maryland, from January 2017 to January 2018. Participants were approached during a routine prenatal visit and enrolled according to predefined inclusion criteria: pregnant at the time of encounter (predetermined by clinic staff); age 18 years or older; able to speak and understand English sufficiently to provide informed consent; and natural hair length at least 3 cm to allow for hair drug testing. All participants provided informed consent and signed a Health Insurance Portability and Accountability Act authorization, to allow access to electronic health records.

All patients entering the clinical sites for prenatal appointments were approached by research staff at check-in and asked to read a brief description of the study to determine interest in participating. If a patient expressed interest, research staff escorted her from the waiting area to a private room, further described the study, and determined whether she met all eligibility criteria. If eligibility criteria were met, informed consent and Health Insurance Portability and Accountability Act authorization were obtained. The research visit took 20–30 minutes on average. Enrolled participants were compensated a total of \$75 for their time (\$50 for first visit and \$25 for 1-week telephone follow-up) using a reloadable gift card.

We conducted this study in accordance with STARD (Standards for Reporting of Diagnostic Accuracy) criteria.¹² As such, all participants were administered the three index tests (4P's Plus, NIDA Quick Screen-ASSIST and SURP-P, in a randomized order) and reference standard (urine and hair drug testing) at the first research visit. The index tests were administered verbally by the research assistants with participants providing verbal answers. Urine samples were collected before the index tests were administered, but results were not available to either the research assistant or the participant until completion of index tests. Hair samples were collected during the same session in which index tests were administered and shipped immediately for reference testing, with results available only after 48 hours. Urine results were shared with participants at the end of the baseline visit, and hair results were shared by telephone within 48 hours of receipt by research staff. Participants who screened positive from screening tools or biological testing were encouraged to speak with their physician about substance use and were offered educational materials and referrals. To assess test–retest reliability of the index tests, we examined the results of repeated screening tool administrations (in randomized order) 1 week apart, with the second administration occurring via telephone, and conducted correlation analysis. To our knowledge, none of the screening tools have been validated for administration over the telephone. Study protocol and methodology are detailed further in a separate report.¹³

The WHO guidelines for identifying and managing substance use during pregnancy⁴ reference 13 validated screening instruments for substance use, but, of those listed, eight assess alcohol only. Of the remaining five, one is an inpatient-only measure and one is a 200-item measure. Three possible brief measures emerged that screen for more than one substance among pregnant women.

The 4P's Plus has been previously validated in a sample of pregnant women.⁸ The 4P's Plus adaptation used in this study consists of seven questions. If there was an affirmative response to any of the latter four questions, the screen was considered "positive" and follow-up questions were asked about past-month quantity of use.

The NIDA Quick Screen-ASSIST is a two-part screening tool. The NIDA Quick Screen⁹ consists of one stem question and then assesses use of: 1) alcohol, 2) tobacco products, 3) prescription drugs for non-medical reasons, and 4) illegal drugs. Only if a participant endorsed use of prescription drugs for nonmedical reasons or illegal drugs in the past year did the interviewer proceed to the ASSIST (items 2–7). For purposes of validation, both the Quick Screen and ASSIST were given to all participants to complete (Table 1). Responses to the ASSIST were summed to create a substance involvement score for each substance. Each substance involvement score was classified using NIDA's classifications as: lower risk (scores 0–3), moderate risk (scores 4–26), or high risk (scores 27 or higher). For validation purposes, moderate and high risk were considered "positive" screens.

The SURP-P¹⁰ consists of three items (Table 1). Scoring involved classifying the number of alcoholic drinks consumed in the month before pregnancy as none compared with any, and then counting the total number of other affirmative items. Negative responses for all items yielded an individual to be considered low-risk, one affirmative response yielded an individual to be considered moderate risk, and two or three affirmative responses yielded an individual to be considered high-risk for substance use (not just alcohol and marijuana). Both moderate- and high-risk classifications were considered a priori to be a screen "positive."

To determine the validity of each screening tool, we used urine and hair testing. Urine testing was used to validate whether a positive screen was indicative of current substance use, which is the primary purpose of the screening tools. It is possible that a participant may not have used substances in the past week but used in the past 3 months. This is a strong possibility in a population of pregnant women who often discontinue use on learning of their pregnancy or as pregnancy progresses. In this case, urine would not validate a positive screen, but hair testing would.

Table 1. Pregnancy Drug Screening Tools*

Screening Tool	Questions
NIDA Quick Screen-ASSIST	
Quick Screen [†]	<ol style="list-style-type: none"> 1. In the past year, how often have you used the following? <ol style="list-style-type: none"> a. Five or more alcohol drinks in a day for men or 4 or more alcohol drinks in a day for women b. Tobacco products c. Prescription drugs for nonmedical reasons d. Illegal drugs
ASSIST [‡]	<ol style="list-style-type: none"> 1. In your lifetime, which of the following substances have you used? (response options of yes or no) 2. In the past 3 mo, how often have you used the substances you mentioned? (response options of never, once or twice, monthly, weekly, and daily or almost daily for items 2–5) 3. In the past 3 mo, how often have you had a strong desire or urge to use (each substance)? 4. (During the past 3 mo, how often has your use of (each substance) led to health, social, legal or financial problems? 5. During the past 3 mo, how often have you failed to do what was normally expected of you because of your use of (each substance)? 6. Has a friend or relative or anyone else ever expressed concern about your use of (each substance)? 7. Have you ever tried to control, cut down or stop using (each substance)? 8. Have you ever used any drug by injection?
SURP-P [§]	<ol style="list-style-type: none"> 1. Have you ever used marijuana? 2. How many alcoholic drinks have you consumed in the month before knowing you were pregnant? 3. Do you feel the need to cut down on your alcohol or drug use?

* 4P's Plus questionnaire not included because it is covered by copyright; the researchers purchased a license to administer to participants.

[†] Response options for each substance are: never, once or twice, monthly, weekly, and daily or almost daily. For purposes of validation, both the Quick Screen and ASSIST were given to all participants to complete.

[‡] Substances assessed are: tobacco products; alcohol; cannabis; cocaine; amphetamine-type stimulants (ATS); sedatives and sleeping pills (benzodiazepines); hallucinogens; inhalants; opioids; and "other" drugs.

[§] Scoring involves classifying the number of alcoholic drinks consumed in the month before pregnancy as none vs any, and then counting the number of affirmative items. Negative responses for all items yields a low-risk individual, one affirmative response yields a moderate risk individual, and two or three affirmative responses yield a high-risk individual.

Although not an indicator of current substance use, we used hair testing to validate the screening tools on less recent substance use. Thus, we used both urine and hair drug testing (combined results) as the reference (gold) standard to capture recent substance use (up to past 90 days).^{14,15} We used the Alere iCup 14-Panel urine multi-drug test to determine the presence of 14 different substances. Hair samples taken at enrollment were sent to a commercial laboratory where screening and confirmatory testing were conducted, thus providing up to a 90-day window of substance use history that allowed us to validate the time frames queried by the three screening tools. Data were collected on all currently prescribed drugs and associated dosage through participants' electronic health records to help distinguish legitimate use from misuse of prescription medications such as buprenorphine, methadone, benzodiazepines and barbiturates. Reference standard test results were not available to assessors at the time of administering index tests.

The primary outcome measures were: the sensitivity, specificity, positive predictive value and negative predictive value of each of the three index tests; and the test-retest reliability of each. Secondary outcome measures were differences in sensitivity, specificity and test-retest reliability for each of the screening tools by age, trimester, and race.

The sample size of 500 was established at the study design phase and determined from a power analysis. The power calculations were based on the

primary aim, which was to conduct validity analyses to determine sensitivity, specificity, and how each screening tool compares to the others and to the reference standard in identifying prescription and illicit drug use. The sample size of 500 participants was based on a one-sample binomial approach, the full methodology, including the sample size derivation, has been published.³ Based on a one-sample binomial approach, with a sample size of 500 participants, as long as no more than 35 individuals test positive in the biological drug tests without a positive screening tool result, we can be 95% confident that the false-negative rate in the population is under 10%. Also, as long as no more than 15 individuals test positive in the urine drug test without a positive survey screen result in the study, we can be 95% confident that the false-negative rate in the population is under 5%. By McNemar's test, if results between any pair of surveys disagreed for at least 15% of study participants, 500 is a sufficient sample size to determine this is significant disagreement. After a preliminary sample size of 500 was chosen, a power analysis was conducted to determine the detectable differences in validity by age, race, and trimester of the enrolled participants. The power of this additional aim with a sample size of 500 was examined. The power of the test of proportions is calculated based on the difference in the proportion of false-negative results in each age group, race, and trimester of pregnancy.

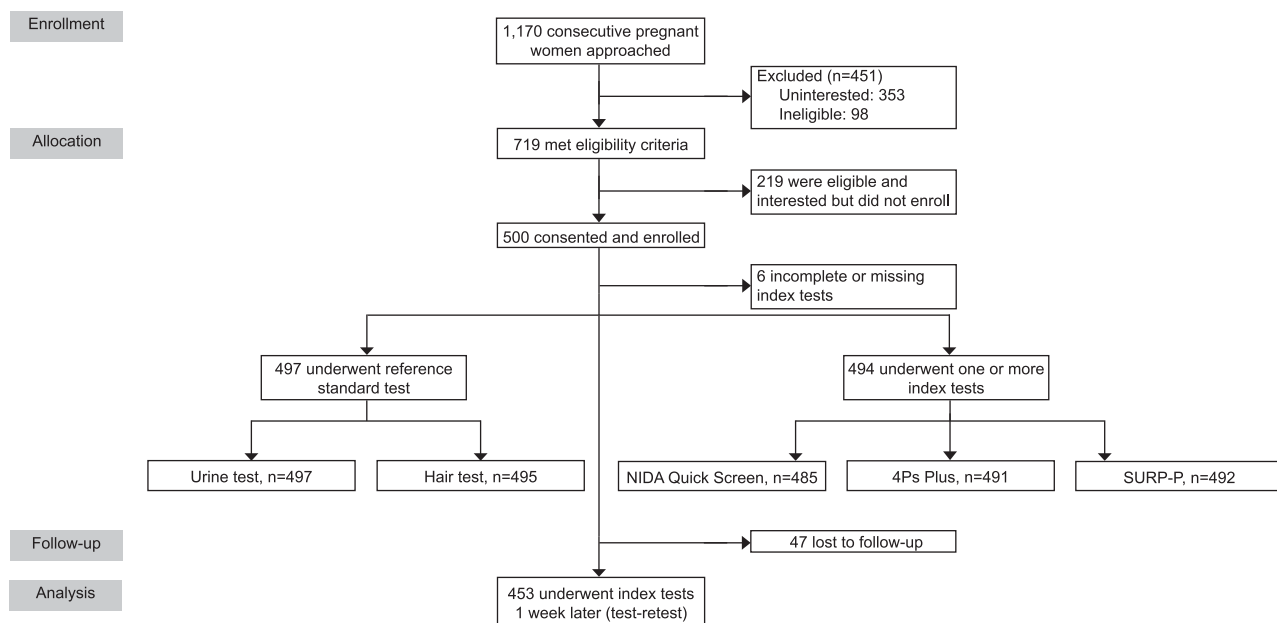


Fig. 1. Flow diagram. SURP-P, Substance Use Risk Profile-Pregnancy scale. Coleman-Cowger. *Prenatal Screening for Substance Use. Obstet Gynecol* 2019.

Descriptive analyses were conducted to show sociodemographic characteristics of the sample. For continuous variables, a one-way analysis of variance model was used to test for a relationship between such variables and trimester; if the necessary assumptions were not met for analyses of variance, a Kruskal-Wallis test was conducted. Chi-square tests for relationships between categorical variables and trimester were conducted. We established sensitivity and specificity for each of the three index tests—4P's Plus, NI-DA Quick Screen-ASSIST, and SURP-P. Sensitivity was calculated as the proportion of persons with a positive reference test who also had positive index tests. Specificity was the proportion of persons with a negative reference test who also had negative index tests. Positive predictive value was the proportion of persons with positive index tests who also had positive reference tests; and negative predictive value was the proportion of persons with negative index tests who

also had negative reference tests. We then calculated test-retest reliability by comparing responses on the index tests with repeat responses obtained 1 week later and provide correlations and phi coefficients for each pair. A phi coefficient of >0.50 was considered acceptable. Invalid or indeterminate reference test (urine or hair) results were excluded from the analysis, as were observations with missing results. Analyses were conducted with Stata version 13.

This study was reviewed and approved by the Institutional Review Boards of the University of Maryland School of Medicine and Battelle Memorial Institute. No adverse events were reported by participants or identified by research staff in connection to this study.

RESULTS

We approached 1,170 pregnant women to participate in our study; 719 (61.5%) were interested and met eligibility criteria; of these, 500 (69.5%) were enrolled

Table 2. Summary Statistics

Variable	Trimester			P*
	1st (n=152)	2nd (n=176)	3rd (n=172)	
Maternal age	27.6±4.92	28.4±5.42	27.5±5.24	.19
No. of previous pregnancies [†]	2 (0–9)	2 (0–10 or more)	2 (0–10 or more)	.41
Race				
African American or black	110 (72.37)	126 (71.59)	116 (67.44)	.68
Caucasian or white	28 (18.42)	37 (21.02)	38 (22.09)	.68
Other or multiracial [‡]	13 (8.55)	12 (6.82)	14 (8.14)	.68
Missing	1 (0.66)	1 (0.57)	4 (2.33)	.68
Education (last grade completed)				
Less than high school	35 (23.03)	28 (15.91)	21 (12.21)	.25
High school graduate	59 (38.82)	72 (40.91)	71 (41.28)	.25
Some college	22 (14.47)	36 (20.45)	28 (16.28)	.25
College graduate	34 (22.37)	38 (21.59)	48 (27.91)	.25
Unavailable	2 (1.32)	2 (1.14)	4 (2.33)	.25
Marital status				
Married [§]	51 (33.55)	66 (37.50)	75 (43.60)	.29
Widowed, divorced, or separated	4 (2.63)	3 (1.70)	6 (3.49)	.29
Never married	96 (63.16)	105 (59.66)	87 (50.58)	.29
Unavailable	1 (0.66)	2 (1.14)	4 (2.33)	.29
Job status				
Working full-time or in the military	72 (47.37)	80 (45.45)	66 (38.37)	.26
Working part-time	19 (12.50)	23 (13.07)	24 (13.95)	.26
Unemployed	59 (38.82)	71 (40.34)	74 (43.02)	.26
Unavailable	2 (1.32)	2 (1.14)	8 (4.65)	.26

Data are mean±SD, median (range), or n (%) unless otherwise specified.

* P values are the results of a one-way analysis of variance model to test for a relationship between age and trimester, a Kruskal-Wallis test for a relationship between number of previous pregnancies and trimester, and χ^2 -square tests for relationships between all remaining variables and trimester.

[†] Number of pregnancies was unavailable for one first-trimester respondent, two second-trimester respondents, and five third-trimester respondents.

[‡] The "Other or multiracial" group includes respondents who chose "Some other group" or more than one category.

[§] The "Married" group includes respondents who were married, living with someone as married, or married but living apart.

^{||} The "Unemployed" group includes respondents who had a job but were unemployed or laid off, those who were not at work for various reasons, full-time homemakers, those in school or training, retired or disabled individuals, and those doing volunteer work.

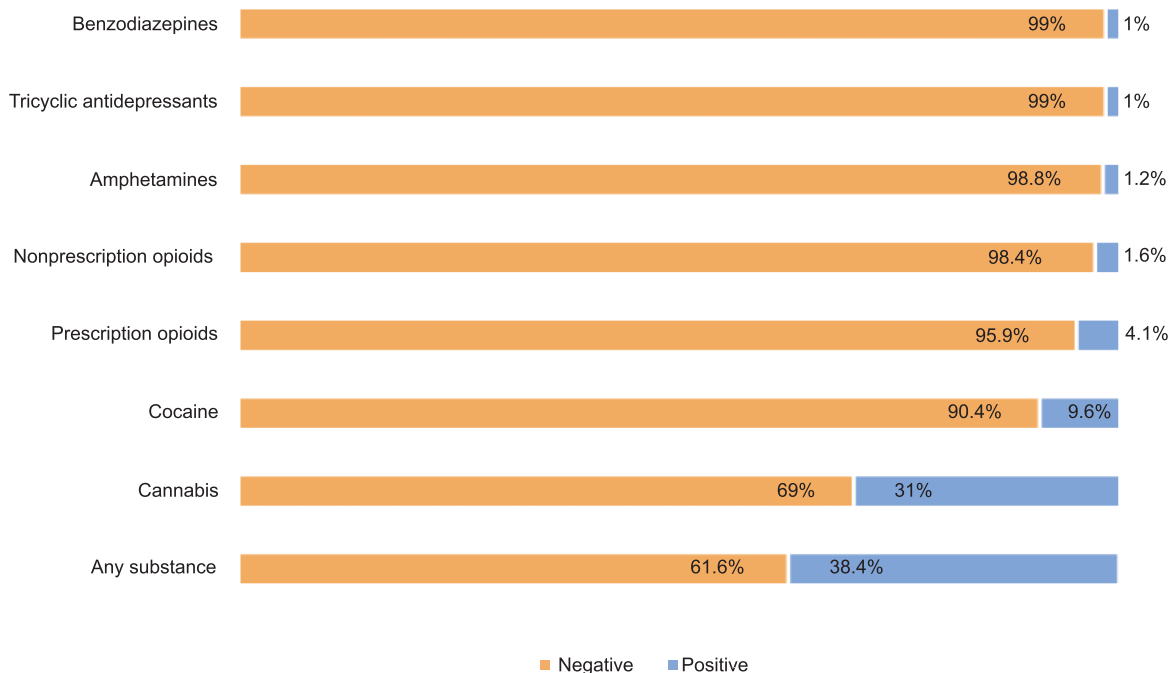


Fig. 2. Prevalence rates by substance.

Coleman-Cowger. *Prenatal Screening for Substance Use. Obstet Gynecol* 2019.

into the study (Fig. 1). Of the enrolled participants, 497 provided biological samples for drug testing (497 urine, 495 hair). A total of 494 participants received at least one of the three index screening tools: 485 were administered the NIDA Quick Screen-ASSIST, 491 were administered the 4P's Plus, and 492 were administered the SURP-P. For test-retest reliability, 453 participants were retested with the three index screening tools, with 47 participants (9.4%) lost to follow-up.

There were 152, 176, and 172 participants in their first, second, and third trimesters, respectively. The distribution of race, education, age, number of previous pregnancies, job status and marital status did not differ across trimesters (Table 2).

Prevalence rates of illicit drug use and prescription drug use as determined by reference standard tests are presented in Figure 2. The most frequently used substance was cannabis, with almost one third of the sample (n=152) testing positive.

Table 3. Validity Indices for the 4P's Plus, NIDA Quick Screen, and SURP-P

	4 P's Plus	NIDA Quick Screen ASSIST	SURP-P
Sensitivity*	91.2 (85.7–95.1)	83.5 (76.8–89.0)	93.1 (88.0–96.5)
Specificity*	28.6 (23.7–33.9)	80.8 (76.0–85.0)	21.0 (16.7–25.9)
Positive predictive value*	39.0 (34.0–44.1)	68.4 (61.3–74.9)	37.0 (32.3–41.9)
Negative predictive value*	86.7 (78.6–92.5)	90.8 (86.8–93.9)	85.9 (76.2–92.7)
Sensitivity [†]	94.7 (88.5–97.4)	85.4 (76.4–89.5)	95.4 (90.7–98.4)
Specificity [†]	28.7 (23.8–33.6)	76.1 (71.4–80.6)	21.1 (17.3–26.1)
Positive predictive value [†]	32.6 (28.9–38.8)	56.4 (50.1–64.4)	30.6 (27.3–36.5)
Negative predictive value [†]	93.6 (85.7–96.7)	93.5 (88.8–95.2)	92.7 (84.8–97.3)
Sensitivity [‡]	90.2 (84.5–93.8)	79.7 (71.2–84.2)	92.4 (87.6–95.8)
Specificity [‡]	29.6 (24.4–35.2)	82.8 (78.1–87.1)	21.8 (17.4–27.2)
Positive predictive value [‡]	44.1 (39.7–50.0)	74.0 (67.8–80.4)	42.0 (38.0–47.9)
Negative predictive value [‡]	83.0 (73.4–88.9)	86.9 (81.3–89.7)	82.3 (72.1–90.0)

Data are % (95% CI).

* Reference standard: hair test results.

[†] Reference standard: urine test results.

[‡] Reference standard: hair and urine test results combined; positive on either urine or hair sample testing.

Table 4. Sensitivity, Specificity, and Test–Retest Analysis by Demographics

Demographic	4 P's Plus		
	Sensitivity	Specificity	Test–Retest*
Age (y)			
18–25	89.9 (81.5–95.6)	40.2 (30.6–52.4)	0.86
26 or older	90.4 (82.2–94.7)	25.2 (19.2–31.3)	0.82
Race			
African American or black	89.8 (84.1–93.9)	36.8 (29.6–44.4)	0.83
Caucasian or white	94.1 (71.3–99.9)	13.3 (6.6–21.7)	0.90
Other or multiracial	75.0 (19.4–99.4)	36.4 (19.8–53.5)	0.76
Trimester			
1st	90.0 (80.7–95.9)	33.3 (23.4–45.4)	0.77
2nd	92.5 (82.3–96.8)	24.3 (16.0–33.6)	0.85
3rd	87.0 (73.7–95.1)	31.9 (23.3–40.9)	0.88

Data are % (95% CI) unless otherwise specified.

* A phi coefficient ≥ 0.50 indicates an acceptable level of correlation.

Table 3 provides results on validity indices of the screening instruments. Using a combination of hair and urine sample drug testing as the reference standard, sensitivity for detecting substance use was higher for the SURP-P and 4P's Plus than the NIDA Quick Screen-ASSIST. Specificity and false-negative rates were highest for the NIDA Quick Screen-ASSIST, followed by the 4P's Plus, then the SURP-P.

Correlation (phi) coefficient for test–retest concordance for the 4P's Plus was 0.84, for NIDA Quick Screen-ASSIST was 0.77, and for SURP-P was 0.79. The mean (SD) number of days from test to retest was 7.7 (1.5). Each test–retest analysis excluded 20 of the total respondents owing to missing data.

Table 4 contains sensitivity, specificity and test–retest reliability by demographic characteristics. Women aged 18–25 years (vs 26 years or older) had significantly lower specificity on the NIDA Quick Screen-ASSIST (70.1% [59.7, 80.0] vs 88.0% [82.8, 92.1]). There were significant differences in specificity by race between non-Hispanic black women and non-Hispanic white women; specificity for the 4P's Plus was 36.8% (29.6, 44.4) and 13.3% (6.6, 21.7), respectively, and specificity for the SURP-P was 29.3% (22.7, 36.7) and 7.8% (3.3, 16.1), respectively. There were no differences in sensitivity or specificity by trimester for any of the three screening tools.

DISCUSSION

To be effective, a screening test must have a high sensitivity to ensure that true-positive results are not missed.¹⁶ Failure to detect and appropriately treat substance use during pregnancy can have long term detrimental effects for both mother and child. In this study validating three self-reported screening tools for

substance use during pregnancy, we found that the SURP-P and 4P's Plus performed similarly with high sensitivity and negative predictive values, making them more ideal screening tests than the NIDA Quick Screen-ASSIST, which had a lower sensitivity with a similar negative predictive value.

High sensitivity often comes at the expense of specificity, which was seen in the performance of these screening tools. The NIDA Quick Screen-ASSIST had the highest specificity, but its low sensitivity makes it less desirable as a screening test. Future studies may consider modifying the language of the NIDA Quick Screen to focus on the past 3 months instead of the past year, given changes in substance use that occur during pregnancy. The SURP-P and 4P's Plus had relatively low specificity. There were differences in performance of the screening tools based on age group, with the NIDA Quick Screen-ASSIST having improved specificity in women older than 25 years. Both SURP-P and 4P's Plus had lower specificity for Caucasian women than for other racial groups. These differences may be related in part to differences in substance preference by subgroups. The three screening tools differ in the extent of substance use they assess. The NIDA Quick Screen-ASSIST assesses behavioral substance use patterns, such as frequency of alcohol, tobacco, and illicit drug use, and it uniquely assesses craving and functional consequences related to substance use. The SURP-P and 4P's Plus do not assess behavioral substance use patterns in such granular detail and do not assess craving or functional consequences of substance use. Furthermore, the SURP-P only inquires about past marijuana and alcohol use and does not assess other substances; the 4P's Plus is less specific

NIDA Quick Screen ASSIST			SURP-P		
Sensitivity	Specificity	Test-Retest*	Sensitivity	Specificity	Test-Retest*
81.0 (69.9–88.3)	70.1 (59.7–80.0)	0.72	91.1 (83.0–96.5)	29.9 (21.0–41.5)	0.76
78.6 (67.9–84.8)	88.0 (82.8–92.1)	0.78	93.3 (86.9–97.3)	18.5 (13.6–24.5)	0.81
80.0 (73.1–85.8)	80.9 (74.3–86.5)	0.74	91.6 (86.3–95.3)	29.3 (22.7–36.7)	0.78
64.7 (38.3–85.8)	87.8 (79.7–94.3)	0.83	100 (80.5–100)	7.8 (3.3–16.1)	0.85
75.0 (19.4–99.4)	78.8 (62.1–91.3)	0.79	100 (39.8–100)	20.6 (8.4–36.9)	0.82
79.7 (68.7–88.6)	84.6 (75.9–92.7)	0.81	91.4 (82.5–96.8)	25.6 (16.6–37.2)	0.86
81.4 (70.3–89.7)	80.4 (70.5–87.2)	0.75	95.5 (88.0–99.1)	18.3 (11.7–27.8)	0.72
71.7 (56.5–84.0)	83.6 (75.8–89.9)	0.76	89.1 (76.4–96.4)	22.4 (15.1–30.8)	0.79

in its assessment of substance use patterns. Each screening tool takes a different approach in assessing substance use but all are intended to screen for multiple substances.

The high false-positive rate needs to be taken into account when recommending these screening tools. The repercussions of a false-positive drug screen cannot be ignored, particularly with stigmatization and the current legal climate regarding pregnant women who use substances, which is punitive in many states.¹⁷ The high negative predictive value of these screening tools means that health care providers can be reasonably assured that a woman who screens negative is not using substances. A positive screen, however, should never be considered diagnostic but instead the nidus for further investigation and initiation of a conversation between the health care provider and patient. The primary purpose of screening tests should be to identify women who may have problematic substance use to provide education, assistance and referral to treatment services to improve their health and pregnancy outcomes.

Of the three screening tests analyzed, two have been previously evaluated in a pregnant population, but only one previous study used biological (urine) confirmatory testing. Although the NIDA Quick Screen-ASSIST has been validated across several populations, it had not previously been validated with pregnant women. Chasnoff et al⁸ found an 87% sensitivity and 76% specificity in the 4P's Plus screening tool in pregnancy. Yonkers et al¹⁰ found a 91% sensitivity and 67% specificity in the SURP-P screening tool for low-risk populations, and a lower sensitivity (57%) and higher specificity (88%) with a high-risk population. Ours is an advancement from these prior studies in that it uses urine and hair screening as the gold-standard confirmatory testing. Using biological

confirmation, the sensitivities were slightly higher in our population than seen in Chasnoff's and Yonkers' validation studies but our specificities were much lower. This may be because the biological testing identified more positives than would have been self-reported in previous studies, and the 4P's Plus and SURP-P do not directly ask about current substance use in as explicit a way as does the NIDA Quick Screen-ASSIST.

Biological screening tests, although considered the gold standard, are not without flaws. Both urine and hair testing have been shown to produce false-positive results and are unable to give information regarding timing or dosage of drug use.¹⁸ Urine tests, although relatively easy to obtain, are subject to variable excretion rates meaning that a negative toxicology test does not necessarily exclude the possibility of recent use, particularly for those drugs with a short half-life. Although hair sampling tests for a longer duration of exposure, its collection is more arduous and not likely to be employed in most obstetric practices.

Although our study has the strength of using two biological tests to confirm use, these tests evaluate for a different group of substances, with hair testing not including benzodiazepines, barbiturates or tricyclic antidepressants, which makes comparison of the two difficult. Neither biological test measured alcohol use, which was not a focus of this study. This may explain why the sensitivity of the screening tools decreased slightly when a combination of hair and urine was taken into account. Drugs are typically eliminated much sooner from the urine than from hair, resulting in different timeframes of use being tested. In our study, 13 participants had inconclusive hair sampling. Although the number of inconclusive hair samples was relatively low in this sample, exclusion of inconclusive tests from analysis may have resulted in

an underestimation of the actual sensitivity of these screening tools.

There were some additional limitations to the study. The study population represents a sample of women willing to enroll in a study regarding substance use screening tools in pregnancy and may inherently be more likely to admit use. The study population was willing to provide biological specimens, and results from biological specimens were reasonably correlated with results from self-report. The confidential nature of the screening tests may have increased the likelihood that women would self-report substance use, although previous studies have shown a relatively high willingness among pregnant women to admit use of substances.¹⁹ Two of the three survey instruments have been studied in pregnancy. The high prevalence of substance use within our population may make it less applicable for lower risk populations. The test–retest reliability (test vs retest administered a week apart) may have been subject to the “practice effect,” a phenomenon in which responses on a questionnaire may be “improved” by prior exposure to the question. However, Marx et al²⁰ compared two retest samples on self-reported quality of life, one 2 days posttest and the other 2 weeks posttest and found no significant differences in test–retest reliability between the two time intervals.

Despite these limitations, this study has several strengths. The use of confirmatory biological markers is an objective comparison to assess the efficacy of a screening tool for prenatal substance use. Comprehensive testing evaluated a large number of substances from both hair and urine. The use of both urine and hair specimens decreases the chance of false-negative results with infrequent use that can be an issue with urine testing alone. The sample size was large, and the population was diverse socioeconomically and well distributed over all trimesters. The test–retest reproducibility of the results was high.

Treatments for prenatal substance use markedly improve outcomes,²¹ lending support to the development and implementation of a screening test according to the WHO Wilson criteria.²² Substance use screening in pregnancy needs to have a lower threshold than in the nonpregnant population because occasional, recreational use likely represents problematic use in a pregnant woman when the same pattern of use may not have qualified as such before conception. Although screening with biological tests such as urine toxicology have utility for confirmatory testing, screening questionnaires are low cost, noninvasive and allow self-report of use, which may provide context and assist in the building of a trusting doctor-

patient relationship which is essential in the treatment of substance use disorders. Both are useful, with some evidence that neither is clearly superior to the other.²³

Our study found that the SURP-P and 4P's Plus were highly sensitive screening tools across all trimesters, races and age groups. The dissemination of a strong and clear recommendation for a clinically useful prescription and illicit drug screening tool for pregnant women is highly significant, relevant for public health, and will likely increase screening, thus providing greater opportunity to intervene with women who may use substances during pregnancy.

REFERENCES

1. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2017 national survey on drug use and health (HHS publication no. SMA 18-5068, NSDUH series H-53). Rockville (MD): Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2018.
2. Center for Behavioral Health Statistics and Quality. 2016 National survey on drug use and health public use file codebook. Rockville (MD): Substance Abuse and Mental Health Services Administration; 2017.
3. Substance Abuse and Mental Health Services Administration. The NSDUH report: substance use among women during pregnancy and following childbirth. Rockville (MD): Substance Abuse and Mental Health Services Administration; 2009.
4. World Health Organization. Guidelines for the identification and management of substance use and substance use disorders in pregnancy. Geneva (Switzerland): World Health Organization; 2014.
5. Opioid use and opioid use disorder in pregnancy. Committee Opinion No. 711. American College of Obstetricians and Gynecologists. *Obstet Gynecol* 2017;130:e81–94.
6. Wright TE, Terplan M, Ondersma SJ, Boyce C, Yonkers K, Chang G, et al. The role of screening, brief intervention, and referral to treatment in the perinatal period. *Am J Obstetrics Gynecol* 2016;215:539–47.
7. Perreira KM, Cortes KE. Race/ethnicity and nativity differences in alcohol and tobacco use during pregnancy. *Am J Public Health* 2006;96:1629–36.
8. Chasnoff I, Wells A, McGourty R, Bailey L. Validation of the 4P's Plus screen for substance use in pregnancy validation of the 4P's Plus. *J Perinatology* 2007;27:744–8.
9. National Institute on Drug Abuse. Resource guide: screening for drug use in general medical settings. Available at: <https://www.drugabuse.gov/publications/resource-guide-screening-drug-use-in-general-medical-settings>. Retrieved August 7, 2018.
10. Yonkers KA, Gotman N, Kershaw T, Forray A, Howell HB, Rounsaville BJ. Screening for prenatal substance use: development of the substance use risk profile-pregnancy scale. *Obstetrics Gynecol* 2010;116:827–33.
11. U.S. Preventive Services Task Force. Final update summary: drug use, illicit: screening. Available at: <https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/drug-use-illicit-screening>. Retrieved June 20, 2018.

12. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; 351:h5527.
13. Coleman-Cowger VH, Oga EA, Peters EN, Trocin K, Koszowski B, Mark K. Comparison and validation of screening tools for substance use in pregnancy: a cross-sectional study conducted in Maryland prenatal clinics. *BMJ Open* 2018;8: e020248.
14. DuPont RL, Baumgartner WA. Drug testing by urine and hair analysis: complementary features and scientific issues. *Forensic Sci Int* 1995;70:63–76.
15. Ledgerwood DM, Goldberger BA, Risk NK, Lewis CE, Price RK. Comparison between self-report and hair analysis of illicit drug use in a community sample of middle-aged men. *Addict behaviors* 2008;33:1131–9.
16. Maxim LD, Niebo R, Utell MJ. Screening tests: a review with examples. *Inhalation Toxicol* 2014;26:811–28.
17. Terplan M, Kennedy-Hendricks A, Chisolm MS. Article commentary: prenatal substance use: exploring assumptions of maternal unfitness. Substance abuse: research and treatment. *Subst Abuse* 2015;9:1–4.
18. Saitman A, Park HD, Fitzgerald RL. False-positive interferences of common urine drug screen immunoassays: a review. *J Anal Toxicol* 2014;38:387–96.
19. Roberts SC, Nuru-Jeter A. Women's perspectives on screening for alcohol and drug use in prenatal care. *Women's Health Issues* 2010;20:193–200.
20. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol* 2003;56:730–5.
21. Kotelchuck M, Cheng ER, Belanoff C, Cabral HJ, Babakhanlou-Chase H, Derrington TM, et al. The prevalence and impact of substance use disorder and treatment on maternal obstetric experiences and birth outcomes among singleton deliveries in Massachusetts. *Matern Child Health J* 2017;21:893–902.
22. Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva (Switzerland): World Health Organization; 1968.
23. Christmas JT, Knisely JS, Dawson KS, Dinsmoor MJ, Weber SE, Schnoll SH. Comparison of questionnaire screening and urine toxicology for detection of pregnancy complicated by substance use. *Obstetrics Gynecol* 1992;80:750–4.

PEER REVIEW HISTORY

Received December 6, 2018. Received in revised form February 7, 2019. Accepted February 14, 2019. Peer reviews and author correspondence are available at <http://links.lww.com/AOG/B352>.

Submit Your Manuscript to *Obstetrics & Gynecology*

To submit your manuscript, visit Editorial Manager™ at
<http://ong.editorialmanager.com>

First-time users: Click the “Register” button on the menu bar and enter the requested information. Upon successful registration, you will be sent an e-mail with instructions to verify your registration.

Authors: Click the “Login” button on the menu bar and log in to the system as “Author.” Then submit your manuscript.

Instructions for authors are available online at
<http://ong.editorialmanager.com>.

rev 1/2019