Date: Jan 17, 2019
To: "Victoria H Coleman-Cowger" [redacted]
From: "The Green Journal" em@greenjournal.org
Subject: Your Submission ONG-18-2286

RE: Manuscript Number ONG-18-2286

Prenatal Screening for Substance Use: Diagnostic Validity of Three Screeners among Pregnant Women

Dear Dr. Coleman-Cowger:

Your manuscript has been reviewed by the Editorial Board and by special expert referees. Although it is judged not acceptable for publication in Obstetrics & Gynecology in its present form, we would be willing to give further consideration to a revised version.

If you wish to consider revising your manuscript, you will first need to study carefully the enclosed reports submitted by the referees and editors. Each point raised requires a response, by either revising your manuscript or making a clear and convincing argument as to why no revision is needed. To facilitate our review, we prefer that the cover letter include the comments made by the reviewers and the editor followed by your response. The revised manuscript should indicate the position of all changes made. We suggest that you use the "track changes" feature in your word processing software to do so (rather than strikethrough or underline formatting).

Your paper will be maintained in active status for 21 days from the date of this letter. If we have not heard from you by Feb 07, 2019, we will assume you wish to withdraw the manuscript from further consideration.

REVIEWER COMMENTS:

Reviewer #1: Thank you for your important work.

In terms of disclosures, since the primary author lists a corporation, there should be some disclosure about how or if that employment impacts upon or relates to the study.

Abstract: well written overall. I would use one sentence in the objectives to tell us why he screens are needed (rather than just doing universal testing?)

Introduction: lines 81-3, while they say that screening should be universal, I assumed this to mean urine or blood screen. What advantage do you see to a questionnaire screen? In other words, make the case as to what this type of screening will add to prenatal care.

Methods: you do a really nice job of describing your methods and process.

Results: you do not need to re-describe values in tables 4 and 5 in the text.

Discussion: Again, I am still not convinced that you have mad the case for me for using these screen rather than doing routine urine screens. Patients with problematic drug use will likely be actively using through pregnancy and will be positive. If someone used cannabis but stopped upon discovering the pregnancy, it seems that this may not be relevant.

Reviewer #2: Doing the type of study reported here is fraught with difficulties and these investigators have overcome many of them in this well designed and performed study. The manuscript is clearly written and understandable by non-substance use specialists and scientists. The conclusions are, for the most part warranted. There are, however, several issues that should be addressed.

1. I'd recommend using he term substance use disorder and avoiding pejorative terms whenever possible.

2. Line 138. It should be noted whether the decision to consider moderate and high risk as positive screens was made a

priori or post hoc. The former would be more rigorous, the latter more susceptible to bias.

3. Line 153ff. The choice of urine and hair tests as the gold standard here has strengths and weaknesses. Each has sensitivity issues, for example. These should be noted in more detail in the discussion. Further, neither are useful for picking up alcohol and indeed no alcohol is reported as detected here (Table 3). The screens, however, include alcohol, which would impact the measures of merit. This should be noted in the manuscript and very likely suggests that a different screener should be used for alcohol use during pregnancy -- VERY IMPORTANT GIVEN THE MULTIPLE DOCUMENTED ANATOMIC AND NEURODEVELOPMENTS BEHAVIORAL ABNORMALITIES REPROTED WITH FETAL ALCOHOL EXPOSURE.

4. Line 165.The power analysis is under-discribed. It would be sensitive to the anticipated prevalence of positive and negative outcome measures. I don't understand the statements re false negative -- surely not a standard way of determining sample size or at least of reporting on it.

5. Line 181. The test-retest reliability is probably overestimated, given learning effects with data being re-collected so close together and perhaps patients wanting to please the research staff (social desirability); this might be noted in ht manuscript

6. Line 215ff. The specificity by race differnces need some discussion later in the paper. It's surey not clear to me why these should have occurred.

7. Line 263. The non-relevance of these findings re alcohol screening in pregnancy should be included here.

8. Line 280. Universal screening is not supported by the findings in this study -- no evidence for enhancing outcomes or reducing costs, especially when you've mentioned negatives like incarceration above. The results would support the use of the two screeners for ilicit drugs in pregnancy.

Reviewer #3:

Line 47 - It is probably unnecessary to give the sample size in the precis.

The introduction is well written, and the authors appear to be knowledgeable about the subject. The goal of this manuscript is to identify quality screening instruments for prenatal substance abuse.

Line 166 - The author's description of the sample size calculation should be more detailed. The authors previously published (BMJ Open. 2018 Feb 17;8(2): e020248 PMID 29455170) this description.

The overall study design is appropriate with the following limitations: 1) Two of the three survey instruments have been studied in pregnancy. Not surprising they performed better. 2) The applaud the efforts to obtain biological specimens. The trade-off is that the sample may not reflect the general population. Interestingly, biological data and patient self-reporting are reasonably correlated.

Table 1 - Do the authors have permission to publish 4P's PLUS questionnaire? It is my understanding that this is proprietary.

Table 3 - This prevalence data could be presented better with a bar graph

STATISTICAL EDITOR COMMENTS:

The Statistical Editor makes the following points that need to be addressed:

Table 3: Need to include CIs for all the prevalence estimates. Footnote should be added clarifying how a (+) was established: by examination of either hair or urine, or by requiring both (Table 4 implies that the results were slightly different. If substantially different, should add a supplemental table showing various criteria for (+) prevalence.

Table 4: The NPV and PPV estimates do not apply to other populations with different prevalence rates of (+) substance. A better metric would be LR(+) and LR(-). Also, given the sample sizes, should round values to nearest .1, not.01 and should include CIs for all estimates.

Table 5: Same comments re: rounding to nearest .1 (phi coefficient could be rounded to nearest .01), including CIs. Should also indicate either in this table or a supplemental table, the sample size for all the demographic subsets for each screening test.

General, lines 47-48: Need to indicate by stats testing the basis for "best" performance. Also, since two screens were cited as being superior to the third, that would make them better, not best, unless there was a stats basis for ordering them from best, better and last. The false negative estimate is redundant, since it is the complement of the sensitivity estimate, but if it is cited, then why is not the false positive estimate also cited? As a screening test, both are important and both have consequences. The false positive rates for 4P's and SURP-P are each quite high (70-80% range), compared to ~ 20% false positive rate for the NIDA test. Another useful metric for the three screening tests would be to state their overall accuracy (true positive + true negative)/(all test results), with CIs. It appears that the usual trade-off between specificity and sensitivity is more favorable to the NIDA screen.

The NIDA screen actually has statistically better AUC and accuracy metrics than did either the 4P's or the SURP-P screens. True, the NIDA screen had more false negatives, but far fewer false positives and generally had better discriminatory metrics than did the other two tests. Applied to the population tested, the NIDA did have higher false negative rates, but the other tests had much worse false positive rates, so the net result depends on the trade-off desired from the screening test.

EDITOR COMMENTS:

1. Thank you for your submission to Obstetrics & Gynecology. In addition to the comments from the reviewers above, you are being sent a notated PDF that contains the Editor's specific comments. Please review and consider the comments in this file prior to submitting your revised manuscript.

The notated PDF is uploaded to this submission's record in Editorial Manager. If you cannot locate the file, contact Randi Zung and she will send it by email - rzung@greenjournal.org.

- Most readers won't know what these two specific tools are before reading your paper. As the precis is supposed to be the "hook" on the table of contents page that draws peopel who use the TOC to determine what papers to read, could you make this more generic. Something like:by comparing results of three different tools to screen prenatal patients for drug abuse to biologic samples, two were found to have high screening characteristics. That's not perfect but do you see what I'm suggesting?

- As I read the tables that describe the different screening tools, it seems that they screen for different things.

- how can a screening tool have diagnostic validity? "Screener" sounds like a person. Would you consider substituting "Screening test or screening tool"?

- how were patients selected?

- were they each given all 3? Have they been validated by phone? On the first administration, how were the given? When were the biologic samples obtained with respect to the screening tools?

- you don't provide all of these in the abstract. you do report the false negative rate, however.

- congratulations to your study team. This is remarkable.

- A false positive rate would also be important. It would be problematic to label a patient as a substance user if she is not. In the manuscript, I hope you report this.

- while assumed, I'm not sure this is known, and you didn't test that in this study.

- ages 15-17 years. please add as appropriate throughout

- ACOG is the American College of Obstetricians and Gynecologists.

- as you've noted 2/3 are validated already. Could you explicate what you mean here by validation?

- STARD guidelines ask you to clarify in methods if data collection was planned before index testing (Propsective) or after (Retrospective). Please include this.

- in abstract you say it was a convenience sample. Was it consecutive or convenience? How did you decide on 500?

- was the order here also randomized?

- could you provide a little introduction about how you chose these (see STARD 10A). It looks like 2 are for any drugs and

SURP-P only for alcohol. Is that correct?

- could you explain that it is a two part test first?

- is this on the Quick Screen or the Assist that response on 2-7 are summed?

- are these NIDA's classifications or yours? If yours, how did you decide this? When did you decide this?

- how so if hair is for long term use? please be clear about ability of biologic samples to screen for EtOH

- spell out

- Please discuss here why you chose these outcomes as opposed to pos and neg predictive values (or including those). in your abstract you also report the false negative rate--if not a primary outcome, perhaps shouldn't be in abstract. As one reviewer noted, you have a fairly high rate of substance abuse in your clinic. Sensitivity and Specificity are important or course and should be independent of prevalence. For others who. might want to consider using these tools, a reporting of predictive values in your setting could help them in assessing in their own populations after implementation what the rates are.

- Based on STARD, please add discussion of 12 ab, 13 ab.

- You don't mention PPV and NPV on lines 162-164.

- were results of biological tests provided to the patients?

- Per STARD, any adverse events?

- This is called a primacy claim (your paper is the first or biggest) and must either be deleted or supported by providing the search terms used, dates, and data bases searched (Medline, Ovid, Pubmed, Google Scholar, etc) in order to substantiate your claim.

- Please include a more thorough comparison of the 3 screening tools in your manuscript. As I read what you've provided here and think about the differences the 4P's plus there is no question about illegal drugs (other than cannibas, not illegal everywhere). The ASSIT tool is more thorough (but tested poorly). The SURP tool only asks specifically about marijuana and Alcohol and the 3rd tool is very general.

- is this based on the screening test results only or on the biologic testing?

2. The Editors of Obstetrics & Gynecology are seeking to increase transparency around its peer-review process, in line with efforts to do so in international biomedical peer review publishing. If your article is accepted, we will be posting this revision letter as supplemental digital content to the published article online. Additionally, unless you choose to opt out, we will also be including your point-by-point response to the revision letter, as well as subsequent author queries. If you opt out of including your response, only the revision letter will be posted. Please reply to this letter with one of two responses:
1. OPT-IN: Yes, please publish my response letter and subsequent email correspondence related to author queries.
2. OPT-OUT: No, please do not publish my response letter and subsequent email correspondence related to author queries.

3. As of December 17, 2018, Obstetrics & Gynecology has implemented an "electronic Copyright Transfer Agreement" (eCTA) and will no longer be collecting author agreement forms. When you are ready to revise your manuscript, you will be prompted in Editorial Manager (EM) to click on "Revise Submission." Doing so will launch the resubmission process, and you will be walked through the various questions that comprise the eCTA. Each of your coauthors will receive an email from the system requesting that they review and electronically sign the eCTA.

Any author agreement forms previously submitted will be superseded by the eCTA. During the resubmission process, you are welcome to remove these PDFs from EM. However, if you prefer, we can remove them for you after submission.

4. Standard obstetric and gynecology data definitions have been developed through the reVITALize initiative, which was convened by the American College of Obstetricians and Gynecologists and the members of the Women's Health Registry Alliance. Obstetrics & Gynecology has adopted the use of the reVITALize definitions. Please access the obstetric and gynecology data definitions at https://www.acog.org/About-ACOG/ACOG-Departments/Patient-Safety-and-Quality-Improvement/reVITALize. If use of the reVITALize definitions is problematic, please discuss this in your point-by-point response to this letter.

5. Because of space limitations, it is important that your revised manuscript adhere to the following length restrictions by manuscript type: Original Research reports should not exceed 26 typed, double-spaced pages (6,500 words). Stated page limits include all numbered pages in a manuscript (i.e., title page, précis, abstract, text, references, tables, boxes, figure legends, and print appendixes) but exclude references.

6. Specific rules govern the use of acknowledgments in the journal. Please note the following guidelines:

* All financial support of the study must be acknowledged.
* Any and all manuscript preparation assistance, including but not limited to topic development, data collection, analysis, writing, or editorial assistance, must be disclosed in the acknowledgments. Such acknowledgments must identify the entities that provided and paid for this assistance, whether directly or indirectly.
* All persons who contributed to the work reported in the manuscript, but not sufficiently to be authors, must be acknowledged. Written permission must be obtained from all individuals named in the acknowledgments, as readers may infer their endorsement of the data and conclusions. Please note that your response in the journal's electronic author form verifies that permission has been obtained from all named persons.
* If all or part of the paper was presented at the Annual Clinical and Scientific Meeting of the American College of Obstetricians and Gynecologists or at any other organizational meeting, that presentation should be noted (include the exact dates and location of the meeting).

7. Provide a précis on the second page, for use in the Table of Contents. The précis is a single sentence of no more than 25 words that states the conclusion(s) of the report (ie, the bottom line). The précis should be similar to the abstract's conclusion. Do not use commercial names, abbreviations, or acronyms in the précis. Please avoid phrases like "This paper presents" or "This case presents."

8. The most common deficiency in revised manuscripts involves the abstract. Be sure there are no inconsistencies between the Abstract and the manuscript, and that the Abstract has a clear conclusion statement based on the results found in the paper. Make sure that the abstract does not contain information that does not appear in the body text. If you submit a revision, please check the abstract carefully.

In addition, the abstract length should follow journal guidelines. The word limits for different article types are as follows: Original Research articles, 300 words. Please provide a word count.

9. Only standard abbreviations and acronyms are allowed. A selected list is available online at http://edmgr.ovid.com/ong/accounts/abbreviations.pdf. Abbreviations and acronyms cannot be used in the title or précis. Abbreviations and acronyms must be spelled out the first time they are used in the abstract and again in the body of the manuscript.

10. The journal does not use the virgule symbol (/) in sentences with words. Please rephrase your text to avoid using "and/or," or similar constructions throughout the text. You may retain this symbol if you are using it to express data or a measurement.

11. Line 246: We discourage claims of first reports since they are often difficult to prove. How do you know this is the first report? If this is based on a systematic search of the literature, that search should be described in the text (search engine, search terms, date range of search, and languages encompassed by the search). If on the other hand, it is not based on a systematic search but only on your level of awareness, it is not a claim we permit.

12. Please review the journal's Table Checklist to make sure that your tables conform to journal style. The Table Checklist is available online here: http://edmgr.ovid.com/ong/accounts/table_checklist.pdf.

13. The American College of Obstetricians and Gynecologists' (ACOG) documents are frequently updated. These documents may be withdrawn and replaced with newer, revised versions. If you cite ACOG documents in your manuscript, be sure the reference you are citing is still current and available. If the reference you are citing has been updated (ie, replaced by a newer version), please ensure that the new version supports whatever statement you are making in your manuscript and then update your reference list accordingly (exceptions could include manuscripts that address items of historical interest). If the reference you are citing has been withdrawn with no clear replacement, please contact the editorial office for assistance (obgyn@greenjournal.org). In most cases, if an ACOG document has been withdrawn, it should not be referenced in your manuscript (exceptions could include manuscripts that address items of historical interest). All ACOG documents (eg, Committee Opinions and Practice Bulletins) may be found via the Clinical Guidance & Publications page at https://www.acog.org/Clinical-Guidance-and-Publications/Search-Clinical-Guidance.

14. Figure 1: Please upload as a separate figure file on Editorial Manager. Additionally, were the 6 with incomplete or missing index tests not excluded (500–6–47 does not equal 453).

15. Authors whose manuscripts have been accepted for publication have the option to pay an article processing charge and publish open access. With this choice, articles are made freely available online immediately upon publication. An information sheet is available at http://links.lww.com/LWW-ES/A48. The cost for publishing an article as open access can be found at http://edmgr.ovid.com/acd/accounts/ifauth.htm.

Please note that if your article is accepted, you will receive an email from the editorial office asking you to choose a publication route (traditional or open access). Please keep an eye out for that future email and be sure to respond to it promptly.

16. If you choose to revise your manuscript, please submit your revision via Editorial Manager for Obstetrics & Gynecology

at http://ong.editorialmanager.com. It is essential that your cover letter list point-by-point the changes made in response to each criticism. Also, please save and submit your manuscript in a word processing format such as Microsoft Word.

If you submit a revision, we will assume that it has been developed in consultation with your co-authors and that each author has given approval to the final form of the revision.

Again, your paper will be maintained in active status for 21 days from the date of this letter. If we have not heard from you by Feb 07, 2019, we will assume you wish to withdraw the manuscript from further consideration.

Sincerely,

Nancy C. Chescheir, MD
Editor-in-Chief

2017 IMPACT FACTOR: 4.982
2017 IMPACT FACTOR RANKING: 5th out of 82 ob/gyn journals

_____

Emmes

February 7, 2019

Nancy Chescheir, MD, Editor
Obstetrics & Gynecology

Dear Dr. Chescheir:

My co-authors and I are grateful for the opportunity to submit a revised version of our manuscript, "Prenatal Screening for Substance Use: Comparing Accuracy of Three Screening Tools among Pregnant Women" for continued consideration as Original Research in *Obstetrics & Gynecology*. We were encouraged by the reviewers' recognition of the importance of our work and appreciated the detailed feedback offered for improving upon our original manuscript. We have addressed the reviewers' and editors' concerns point-by-point.

Reviewer #1:
"In terms of disclosures, since the primary author lists a corporation, there should be some disclosure about how or if that employment impacts upon or relates to the study."

Dr. Coleman-Cowger is employed by and receives a salary from The Emmes Corporation. However, this employment does not impact upon or relate to the study, which is funded by a grant from NIDA. The grant was originally awarded and data collection was completed while Dr. Coleman-Cowger was at Battelle.

"Abstract: well written overall. I would use one sentence in the objectives to tell us why the screens are needed (rather than just doing universal testing?)"

Thank you for this comment. We agree that it is important for readers to understand the difference between screening tools and biologic testing and have added a sentence to the Objectives. We have also added the following information to the introduction on lines 86-87: "Although many providers use biologic testing to determine use, a positive urine toxicology does not provide any context regarding temporality of use or indications of problematic use."

"Introduction: lines 81-3, while they say that screening should be universal, I assumed this to mean urine or blood screen. What advantage do you see to a questionnaire screen? In other words, make the case as to what this type of screening will add to prenatal care."

We agree with your statement and thank you for this feedback. The following statement has been added to lines 86-87 in the introduction: "Although many providers use biologic testing to determine use, a positive urine toxicology does not provide any context regarding temporality of use or indications of problematic use."

"Methods: you do a really nice job of describing your methods and process."

Thank you.

"Results: you do not need to re-describe values in tables 4 and 5 in the text."

We have edited the text in the Results section so as not to re-state the values from tables 4 and 5 in the text, except for race differences which we felt was clearer to keep in text.

"Discussion: Again, I am still not convinced that you have made the case for me for using these screens rather than doing routine urine screens. Patients with problematic drug use will likely be actively using through pregnancy and will be positive. If someone used cannabis but stopped upon discovering the pregnancy, it seems that this may not be relevant."

The following sentence was added to lines 259-262: "Urine tests, although relatively easy to obtain, are subject to variable excretion rates meaning that a negative toxicology test does not necessarily preclude 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." We have also added the following sentence to lines 290-292: "Although screening with biologic tests such as urine toxicology have utility for confirmatory testing, screening questionnaires are low cost, non-invasive 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."

Reviewer #2:
"I'd recommend using the term substance use disorder and avoiding pejorative terms whenever possible."

In this manuscript, we utilize the terminology "illicit drug use" and "prescription drug misuse" to align with that used by the National Survey on Drug Use and Health, but otherwise use "substance use" throughout. We do not use "substance use disorder" as the screening tools are not diagnostic and are screening for substance use only.

"Line 138. It should be noted whether the decision to consider moderate and high risk as positive screens was made a priori or post hoc. The former would be more rigorous, the latter more susceptible to bias."

We have noted in the manuscript that this was an a priori decision.

"Line 153ff. The choice of urine and hair tests as the gold standard here has strengths and weaknesses. Each has sensitivity issues, for example. These should be noted in more detail in the discussion. Further, neither are useful for picking up alcohol and indeed no alcohol is reported as detected here (Table 3). The screens, however, include alcohol, which would impact the measures of merit. This should be noted in the manuscript and very likely suggests

that a different screener should be used for alcohol use during pregnancy -- VERY IMPORTANT GIVEN THE MULTIPLE DOCUMENTED ANATOMIC AND NEURODEVELOPMENTS BEHAVIORAL ABNORMALITIES REPROTED WITH FETAL ALCOHOL EXPOSURE."

We have addressed the strengths and weaknesses of biologic testing in the Discussion. Drug testing with hair and urine samples have limitations in that urine and hair testing can produce false positives and cannot tell timing or dosage of substance use. Hair sample testing does not include benzodiazepines, barbiturates or tricyclic antidepressants, additionally, hair samples allow for detecting substance use for up to 90 days after last use compared to 5-14 days for urine. It is in recognition of these individual limitations that we used a combination of the two, allowing for detection of recent substance use, and up to 90 days afterwards. We could not determine from this study the accuracy of the selected screeners in detecting prenatal alcohol use, as ethanol was not measured via biologic testing. The purpose of this study was specific to illicit and prescription drugs and no conclusions can be drawn about the accuracy of the screening tools for alcohol.

"Line 165.The power analysis is under-described. It would be sensitive to the anticipated prevalence of positive and negative outcome measures. I don't understand the statements re false negative -- surely not a standard way of determining sample size or at least of reporting on it."

We have addressed this and expanded the description of the sample size calculations.

"Line 181. The test-retest reliability is probably overestimated, given learning effects with data being re-collected so close together and perhaps patients wanting to please the research staff (social desirability); this might be noted in the manuscript."

We have addressed this concern by expanding our discussion.

"Line 215ff. The specificity by race differences need some discussion later in the paper. It's not clear to me why these should have occurred."

We speak more to the observed race differences in the Discussion, noting that these may be due to differences in substance use by subgroup and differences in the screening tools and how they assess substance use.

"Line 263. The non-relevance of these findings re alcohol screening in pregnancy should be included here."

We acknowledge in the Discussion that this study did not focus on alcohol screening.

"Line 280. Universal screening is not supported by the findings in this study -- no evidence for enhancing outcomes or reducing costs, especially when you've mentioned negatives like

incarceration above. The results would support the use of the two screeners for illicit drugs in pregnancy."

Thank you for this comment. We have removed this line from the discussion/conclusions.

Reviewer #3:
"Line 47 - It is probably unnecessary to give the sample size in the precis."

We have removed the sample size from the precis.

"Line 166 - The author's description of the sample size calculation should be more detailed. The authors previously published (BMJ Open. 2018 Feb 17;8(2): e020248 PMID 29455170) this description."

We have expanded the description of the sample size calculations.

"The overall study design is appropriate with the following limitations: 1) Two of the three survey instruments have been studied in pregnancy. Not surprising they performed better. 2) The applaud the efforts to obtain biological specimens. The trade-off is that the sample may not reflect the general population. Interestingly, biological data and patient self-reporting are reasonably correlated."

We have added these limitations to the Discussion section.

"Table 1 - Do the authors have permission to publish 4P's PLUS questionnaire? It is my understanding that this is proprietary."

We have removed reference to the specific questions from the 4P's Plus and stated copyright limitations in the footnotes.

"Table 3 - This prevalence data could be presented better with a bar graph."

We appreciate the reviewer's suggestion and have removed Table 3. We now present the data from that table in a bar graph (Figure 2).

Statistical Editor:
"Table 3: Need to include CIs for all the prevalence estimates. Footnote should be added clarifying how a (+) was established: by examination of either hair or urine, or by requiring both (Table 4 implies that the results were slightly different. If substantially different, should add a supplemental table showing various criteria for (+) prevalence."

We have addressed this in the manuscript.

"Table 4: The NPV and PPV estimates do not apply to other populations with different prevalence rates of (+) substance. A better metric would be LR(+) and LR(-). Also, given the sample sizes, should round values to nearest .1, not.01 and should include CIs for all estimates."

We have included confidence intervals and corrected the decimal places.

"Table 5: Same comments re: rounding to nearest .1 (phi coefficient could be rounded to nearest .01), including CIs. Should also indicate either in this table or a supplemental table, the sample size for all the demographic subsets for each screening test."

We have addressed this in the manuscript.

"General, lines 47-48: Need to indicate by stats testing the basis for "best" performance. Also, since two screens were cited as being superior to the third, that would make them better, not best, unless there was a stats basis for ordering them from best, better and last. The false negative estimate is redundant, since it is the complement of the sensitivity estimate, but if it is cited, then why is not the false positive estimate also cited? As a screening test, both are important and both have consequences. The false positive rates for 4P's and SURP-P are each quite high (70-80% range), compared to ~ 20% false positive rate for the NIDA test. Another useful metric for the three screening tests would be to state their overall accuracy (true positive + true negative)/(all test results), with CIs. It appears that the usual trade-off between specificity and sensitivity is more favorable to the NIDA screen."

We have removed reference to "better" or "best" as we include all the indices for the reader to make an informed judgment about each screening tool's usefulness.

"The NIDA screen actually has statistically better AUC and accuracy metrics than did either the 4P's or the SURP-P screens. True, the NIDA screen had more false negatives, but far fewer false positives and generally had better discriminatory metrics than did the other two tests. Applied to the population tested, the NIDA did have higher false negative rates, but the other tests had much worse false positive rates, so the net result depends on the trade-off desired from the screening test."

We agree and leave the trade-off decisions up to the reader.

<u>Editor</u>:
"Most readers won't know what these two specific tools are before reading your paper. As the precis is supposed to be the "hook" on the table of contents page that draws people who use the TOC to determine what papers to read, could you make this more generic…"

We have edited the precis to be less specific and capture a reader's attention.

"As I read the tables that describe the different screening tools, it seems that they screen for different things."

We have included a more thorough comparison of the 3 screening tools, and what they screen for, in the Discussion section.

“how can a screening tool have diagnostic validity?”

We have changed the wording for clarity to simply note “validity.” Given that we are examining the accuracy of screening tools in identifying substance use rather than a diagnosis of substance use disorder, we recognize that the terminology diagnostic validity could be confusing. We also state in the discussion that a positive screen is not intended to be diagnostic.

"Screener" sounds like a person. Would you consider substituting "Screening test or screening tool"?

We have edited the text throughout to “screening tool” instead of “screener.”

“how were patients selected?”

Patient selection is detailed in the materials and methods section of the paper; however, the abstract has been corrected to state the sample as “consecutive” rather than “convenience.”

“were they each given all 3? Have they been validated by phone? On the first administration, how were the given? When were the biologic samples obtained with respect to the screening tools?”

Each participant was given all 3 screening tools in a randomized order. The tools have not been validated for telephone administration, and this is now noted in the text, along with answers to each of the editor’s subsequent questions.

“you don't provide all of these in the abstract. you do report the false negative rate, however.”

We generally agree with the reviewer’s sentiments on this. As such, we have stuck to reporting only sensitivity and specificity in the abstract. We have also taken out false negative rates, since these are really just the inverse of sensitivity.

“A false positive rate would also be important. It would be problematic to label a patient as a substance user if she is not. In the manuscript, I hope you report this.”

Thank you for this comment. We completely agree. We feel that this issue was addressed in the paragraph in the discussion that begins “The high false positive rate needs to be taken into account when recommending these screeners…”

“while assumed, I'm not sure this is known, and you didn't test that in this study.”

We have removed this part of the sentence from the abstract.

"ages 15-17 years. please add as appropriate throughout"

We have now added "years" throughout.

"ACOG is the American College of Obstetricians and Gynecologists."

This correction has been made.

"as you've noted 2/3 are validated already. Could you explicate what you mean here by validation?"

Validation here refers to validation by hair and urine biologic testing confirmation. Previously validated studies were done by interviews or urine only.

"STARD guidelines ask you to clarify in methods if data collection was planned before index testing (Prospective) or after (Retrospective). Please include this."

We now state that data collection was prospective.

"in abstract you say it was a convenience sample. Was it consecutive or convenience? How did you decide on 500?"

It was consecutive. This has been corrected in the abstract. Our sample of 500 was decided upon prior to study start and was based on preliminary clinic data and power calculations.

"was the order here also randomized?"

Yes, the order of index tests was randomized in the follow-up as well. We have added this information to the manuscript.

"could you provide a little introduction about how you chose these (see STARD 10A). It looks like 2 are for any drugs and SURP-P only for alcohol. Is that correct?"

Introductory text has been added to the Index Tests section to clarify how the three screening tools were chosen. All are intended to be screening tools for substance use, including the SURP-P (though the questions only inquire about alcohol and marijuana use).

"could you explain that it is a two part test first?"

We have added a sentence explaining that this is a two-part screening tool.

"is this on the Quick Screen or the Assist that response on 2-7 are summed?"

The ASSIST contains items 2-7. We have modified the wording to make this more clear in the text.

"are these NIDA's classifications or yours? If yours, how did you decide this? When did you decide this?"

These are NIDA's classifications.

"how so if hair is for long term use? please be clear about ability of biologic samples to screen for EtOH"

The text was missing the word "less" so that it read incorrectly. It has been corrected. We are now clear in our Limitations section that alcohol use was not biochemically verified as our tests did not measure EtOH.

"spell out"

We have edited to spell out electronic health records.

"Please discuss here why you chose these outcomes as opposed to pos and neg predictive values (or including those). in your abstract you also report the false negative rate--if not a primary outcome, perhaps shouldn't be in abstract. As one reviewer noted, you have a fairly high rate of substance abuse in your clinic. Sensitivity and Specificity are important or course and should be independent of prevalence. For others who. might want to consider using these tools, a reporting of predictive values in your setting could help them in assessing in their own populations after implementation what the rates are."

"Based on STARD, please add discussion of 12 ab, 13 ab."

This has been addressed within the manuscript.

"You don't mention PPV and NPV on lines 162-164."

We have added NPV and PPV to this paragraph. We thank the reviewer for this comment.

"were results of biological tests provided to the patients?"

Results of urine tests were provided to patients at the in-person baseline visit; results of hair tests were provided by phone within 48 hours of receipt by research staff. We now include this in the Procedure and Test Methods section.

"Per STARD, any adverse events?"

There were no adverse events and we now state this in the Ethical Considerations section of the text.

"This is called a primacy claim (your paper is the first or biggest) and must either be deleted or supported by providing the search terms used, dates, and data bases searched (Medline, Ovid, Pubmed, Google Scholar, etc) in order to substantiate your claim."

We have deleted the primacy claim, and revised language to show that our study is an advancement from prior studies that did not utilize both hair and urine testing.

"Please include a more thorough comparison of the 3 screening tools in your manuscript. As I read what you've provided here and think about the differences the 4P's plus there is no question about illegal drugs (other than cannabis, not illegal everywhere). The ASSIST tool is more thorough (but tested poorly). The SURP tool only asks specifically about marijuana and Alcohol and the 3rd tool is very general."

We have included a more thorough comparison of the 3 screening tools in the Discussion section.

"is this based on the screening test results only or on the biologic testing?"

"The word limits for different article types are as follows: Original Research articles, 300 words. Please provide a word count."

We have included a word count immediately following our abstract (296 words).

"The journal does not use the virgule symbol (/) in sentences with words. Please rephrase your text to avoid using "and/or," or similar constructions throughout the text. You may retain this symbol if you are using it to express data or a measurement."

We have edited the body of the manuscript to remove the virgule symbol.

"Line 246: We discourage claims of first reports since they are often difficult to prove. How do you know this is the first report? If this is based on a systematic search of the literature, that search should be described in the text (search engine, search terms, date range of search, and languages encompassed by the search). If on the other hand, it is not based on a systematic search but only on your level of awareness, it is not a claim we permit."

We have deleted the primacy claim, and revised language to show that our study is an advancement from prior studies that did not utilize both hair and urine testing.

"Please review the journal's Table Checklist to make sure that your tables conform to journal style. The Table Checklist is available online here:
http://edmgr.ovid.com/ong/accounts/table_checklist.pdf."

We have reviewed the Table Checklist and have modified Table 4 given the guidance listed about not having different headings mid-title.

"The American College of Obstetricians and Gynecologists' (ACOG) documents are frequently updated. These documents may be withdrawn and replaced with newer, revised versions. If you cite ACOG documents in your manuscript, be sure the reference you are citing is still current and available. If the reference you are citing has been updated (ie, replaced by a newer version), please ensure that the new version supports whatever statement you are making in your manuscript and then update your reference list accordingly (exceptions could include manuscripts that address items of historical interest)..."

We have checked all ACOG documents cited to ensure that the most up-to-date versions are referenced.

"Figure 1: Please upload as a separate figure file on Editorial Manager. Additionally, were the 6 with incomplete or missing index tests not excluded (500–6–47 does not equal 453)."

We have uploaded Figure 1 as a separate file. The six participants with incomplete or missing index tests were not necessarily excluded from the study if they completed follow-up index tests.

Thank you again for the opportunity to address reviewer concerns and strengthen our manuscript. We wish to "opt-in" to have our response to reviewer concerns published.

Sincerely,

Victoria H. Coleman-Cowger, Ph.D.
[redacted]

From: [redacted]
To: Randi Zung
Subject: Re: Your Revised Manuscript 18-2286R1
Date: Friday, February 15, 2019 12:37:27 AM

Dear Randi,

I have reviewed and have no further edits on v3.

Thanks so much,
Victoria

Victoria H. Coleman-Cowger, Ph.D.
[redacted]

From: Randi Zung <rzung@greenjournal.org>
Sent: Thursday, February 14, 2019 2:51 PM
To: Victoria Coleman Cowger
Subject: RE: Your Revised Manuscript 18-2286R1

Dear Dr. Coleman-Cowger:

The Editors have discussed your manuscript.

Regarding the title, we are going to revert back to, “Accuracy of Three Screening Tools for Prenatal Substance Use.”

In your Abstract-Objective, the Editors would like you to avoid using “determine.” I have replaced this with “evaluate the” in the attached file (v3). If there is any similar phrasing that you need to edit for consistency, please make those changes. This will be your final opportunity to make substantive edits to the text.

If you do not have any further changes, please let me know that we are okay to proceed with v3. If you do have changes to make, please send me your updated file.

Thank you,
Randi

From: Victoria Coleman Cowger [redacted]
Sent: Wednesday, February 13, 2019 12:06 PM
To: Randi Zung <RZung@greenjournal.org>
Subject: RE: Your Revised Manuscript 18-2286R1 - Correction

Hi Randi,

Either "Problematic" or "Hazardous" Substance Use would be fine to move forward with. I don't think the title needs both. Whatever the editors are more comfortable with would be fine with me. My preference, just to note it, would still be Substance Use only since that is actually what the screeners are screening for and the tone of Problematic or Hazardous could be misconstrued/difficult to define.

Thank you,
Victoria

From: Randi Zung <RZung@greenjournal.org>
Sent: Wednesday, February 13, 2019 10:45 AM
To: Victoria Coleman Cowger [redacted]
Subject: RE: Your Revised Manuscript 18-2286R1 - Correction

Dear Dr. Coleman-Cowger:

I have spoken to the Manuscript Editor. Would you be okay with proceeding with: "Accuracy of Three Screening Tools for Problematic or Hazardous Substance Use Among Pregnant Women"?

The Editors will be discussing your manuscript tomorrow so I am trying to provide them with the most accurate title possible for their conference.

Thanks,
Randi

From: Victoria Coleman Cowger [redacted]
Sent: Wednesday, February 13, 2019 10:21 AM
To: Randi Zung <RZung@greenjournal.org>
Subject: RE: Your Revised Manuscript 18-2286R1 - Correction

Hi Randi,

I'm very hesitant about this title because the screening tools are not diagnostic and substance use disorder is a diagnosis. I would definitely prefer Problematic or Hazardous Substance Use if at all possible.

From: Randi Zung <RZung@greenjournal.org>
Sent: Wednesday, February 13, 2019 10:15 AM
To: Victoria Coleman Cowger [redacted]
Subject: RE: Your Revised Manuscript 18-2286R1 - Correction

Dear Dr. Coleman-Cowger:

Apologies, the information in my previous email below is not correct. The term was updated to "substance use disorder."

"Accuracy of Three Screening Tools for Prenatal Substance Use Disorder" is the title that we would like to propose.

Please let me know if that will be okay.

Thank you,
Randi

From: Victoria Coleman Cowger <[redacted]
Sent: Wednesday, February 13, 2019 10:13 AM
To: Randi Zung <RZung@greenjournal.org>
Subject: RE: Your Revised Manuscript 18-2286R1

Yes, this is fine.

Thank you,
Victoria

From: Randi Zung <RZung@greenjournal.org>
Sent: Wednesday, February 13, 2019 10:07 AM
To: Victoria Coleman Cowger <[redacted]

Subject: RE: Your Revised Manuscript 18-2286R1

Dear Dr. Coleman-Cowger:

The edits to your title were made to bring it in line with the style used by the American College of Obstetricians and Gynecologists. The College style is to use the phrase “substance abuse,” not “substance use.”

Would you be okay with us reverting the title back to, “Accuracy of Three Screening Tools for Prenatal Substance Abuse”?

Thank you,
Randi

From: Victoria Coleman Cowger <[redacted]
Sent: Tuesday, February 12, 2019 8:57 PM
To: Randi Zung <RZung@greenjournal.org>
Subject: RE: Your Revised Manuscript 18-2286R1

Dear Ms. Zung,

Thank you for the opportunity to address these queries. I have attached a tracked changes version of the manuscript to this email and have responded to the author queries both in the manuscript and below inblue.

All my best,
Victoria

****

Victoria H. Coleman-Cowger, Ph.D.
[redacted]

---------- Forwarded message ---------

From: **Randi Zung** <RZung@greenjournal.org>

Date: Tue, Feb 12, 2019 at 8:29 AM

Subject: Your Revised Manuscript 18-2286R1

To: [redacted]

Dear Dr. Coleman-Cowger:

Your revised manuscript is being reviewed by the Editors. Before a final decision can be made, we need you to address the following queries. Please make the requested changes to the latest version of your manuscript that is attached to this email. **Please track your changes and leave the ones made by the Editorial Office.** Please also note your responses to the author queries in your email message back to me.

1. General: The Editor has made edits to the manuscript using track changes. Please review them to make sure they are correct.

I have reviewed all changes and they are correct, with the exception of one that I rejected related to query 11. Thank you.

2. Title: Do you agree with the edited title?

I modified it slightly to say “use” and not “abuse” but otherwise I agree with the edits.

3. Kathleen E. Trocin will need to complete our electronic Copyright Transfer Agreement, which was sent to her through Editorial Manager

Ms. Trocin has completed the electronic Copyright Transfer Agreement.

4. Abstract-Objective: The objective for the abstract should be a simple "to" statement without background. The second sentence listed here is not needed.

Okay. I agree with edits made.

5. Starting at Line 63 and elsewhere: Please remove the use of the word “screener” and replace throughout with “screening test”.

We had changed many references of screener to screening tool in the prior revision but missed several. We have made the change here and did a search and replace for all other reference to “screeners.”

6. Line 67: Table 3 says 84.5. Which is correct?

The number in the table is correct (84.5) and we have corrected the abstract to reflect this.

7. Line 112: Please clarify. I have a hard time believing that every woman approached was enrolled for a total of 500. The # of enrolled belongs in the results section. There, please tell us how many consecutive patients were approached and how many agreed. Please edit this statement like you have it in the first paragraph of the Results.

We have modified this sentence for clarity and removed number of enrolled participants from this section. The number of approached and enrolled participants is reported in Results.

8. Line 138: By "screened positive" here does that include only the questionnaire screening tests or does it also include the biologic testing? I ask because right before this sentence you are describing handling of the biological testing results.

It includes both. I have added additional clarifying information to this sentence.

9. Line 167: 2 or 3 affirmative responses don't "yield a HR individual". 2 or 3 affirmative responses might result in someone being considered to be at high risk. Same is true for the "yielded a low-risk individual" above.

We appreciate the suggestion to change this language, and we have modified this sentence.

10. Line 203: This might be clearer, if I understand the meaning correctly if you said "If results between any pair of surveys disagreed for at least 15% of study participants,…".

We have changed wording as suggested for clarity.

11. Line 277: I'm a bit confused here. Have copied what you said above about the SURP-P study. There you indicated that "2 or 3 affirmative responses yielded a high risk individual for substance use, not just alcohol and marijuana." You reiterate here that it screens only for MJ and EtOH. So how can you say it identifies people at high risk for use of other drugs? Based on what? Also, the tests.

The SURP-P is intended to be a screening tool for all prenatal substance use, though it only inquires about alcohol and marijuana use. The questions asked are about ever-use of marijuana and alcohol use in the month prior to pregnancy, as the idea is that women may not be as forthcoming about substance use during pregnancy if asked directly. These questions serve as a proxy for current substance use. Here's the article about the development of the SURP-P: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 and gynecology. 2010 Oct;116(4):827.

12. Line 285: I always have trouble with this as I know what you mean, but it's not technically correct. Neither alcohol, nor marijuana in many states, are "illicit" which generally means forbidden by law or rule.

We have removed "illicit" here.

To facilitate the review process, we would appreciate receiving a response within 48 hours.

Best,
Randi Zung

_ _

Randi Zung (Ms.)

Editorial Administrator | *Obstetrics & Gynecology*

The American College of Obstetricians and Gynecologists

409 12th Street, SW

Washington, DC 20024-2188

T: 202-314-2341 | F: 202-479-0830

http://www.greenjournal.org