Socioeconomic status and race as social determinants of health to be considered in clinical use of pre-season vestibular and oculomotor tests for concussion

Jessica Wallace, Phillip Worts, Ryan Moran, Justin Mason, Katherine K. Weise, Mark Swanson, Nicholas Murray

*Corresponding author
Jessica Wallace
University of Alabama – Department of Health Science 270 Kilgore Lane, Capital Hall 2106 Tuscaloosa, AL 35405

1st Editorial decision
08-Jul-2020

Ref.: Ms. No. JCTRes-D-20-00049
Socioeconomic Status and Racial Disparities to be Considered in Clinical Interpretation of
Pre-Season Vestibular and Oculomotor Tests for Concussion
Journal of Clinical and Translational Research

Dear Dr. Wallace,

Reviewers have now commented on your paper. You will see that they are advising that you revise your manuscript. If you are prepared to undertake the work required, I would be pleased to reconsider my decision.

For your guidance, reviewers' comments are appended below.

If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript.Also, please ensure that the track changes function is switched on when implementing the revisions. This enables the reviewers to rapidly verify all changes made.

Your revision is due by Aug 07, 2020.

To submit a revision, go to https://www.editorialmanager.com/jctres/ and log in as an Author. You will see a menu item call Submission Needing Revision. You will find your submission record there.

Yours sincerely

Michal Heger
Editor-in-Chief
Journal of Clinical and Translational Research

Reviewers' comments:

Reviewer #1: This manuscript proposes to investigate an important topic - racial and socioeconomic differences in performance on the VOMS and King Devick instruments for concussion. Unfortunately, the paper diverges from that aim by partly analyzing the data contrary to that aim. In particular, the logistic regression modeling for free lunch (FRL) is backwards in that FRL is not a performance measure. It should be an independent variable (as the aim states) and not the outcome. All of the modeling should involve the performance outcomes - VOMS, Near Point Convergence and K-D. The main focus should be whether race or SES is associated with performance on these measures. The models of FRL confuse the issue and take up much space in the text.
The authors also seem to whittle down their sample size needlessly. Since the outcomes are (or should be) modeled separately, there is no reason to exclude the 99 people who had one of the two outcome measures but not the other. They make a similar exclusion based on race, eliminating an "other" group that also has 99 people. This group is not so small that it would not add to the understanding of race in the modeling. There also seems to be confusion between race and ethnicity (Hispanic). Latinx individuals can be either Black or White and should be included if the question was asked correctly in two parts.
There are also concerns with the modeling beyond the use of FRL as the outcome. The authors continually split the sample and fit individual models by sex and race. This approach seems analogous to a fishing expedition where they are searching for additional statistically significant results. A better approach would be to examine possible interactions in the models for the whole sample in order to determine if stratification is warranted. For example, on page 14, the model for NPC distance showed a significant race effect for those who receive FRL but not those who do not. This may be a power issue not a real finding since the non=FRL group is smaller and the black non-FRL group may be quite small further reducing the power of the comparison.
Minor concerns
There seems to be little point to including the medians in the tables since all but one is zero. A note in the text would suffice. Also, perhaps Tables 1 and 2 could be combined.
The text under Figure 1 draws a conclusion rather than describing the figure. Conclusions should go in the text in the conclusions section.
The "operation definitions" on Page 6 repeats much of what is in the methods on Page 5. It only needs to be there once.
I was surprised that almost half of the people prescribed glasses do not wear them. As someone who has worn glasses most of my life, this is perplexing as to how they function.

Reviewer #2: This manuscript discusses race and SES differences among commonly used vestibulo-ocular measures. Findings indicated that black athletes had higher mean NPC distances and free lunch status was associated with slower K-D times and more provocation on visual motion sensitivity. The authors conclude that race and SES should be considered as modifiers across concussion baseline measures. Specific comments below:

Abstract
Aim
- The term "performance" is used too many times. Consider removing the first reference so that it reads "racial and socioeconomic differences on baseline…"
Methods
- Given the topic of the paper, authors could consider including the percentage of black and FRL participants in addition to the age to provide readers with context.
- Authors use the term SES without previously defining.
- Consider providing analyses used in a brief sentence.
Results
- I would encourage the authors to use more specific terminology when describing directionality of their findings. Specifically, "higher mean NPC distance" is more descriptive than "worse." Same for visual motion sensitivity (e.g., "higher mean provocation") and K-D times (e.g., "slower").
Conclusions
- Authors should define terms such as VMS before abbreviating.
- The point on future research does not seem to fit here and is likely better suited for the discussion section of the paper. Consider removing.

Introduction
The introductory paragraph should be more specific to the paper. The reader should finish the first paragraph and have a sense of the overall topic of the paper (i.e., statement of the problem). Given that this paper does not discuss post-concussion findings, the focus on explaining concussion does not seem to fit. Consider restructuring to focus more on the increased reliance on visual and vestibular measures in concussion management and the need to improve methodology for interpreting those measures.

- I recommend that P4 L15-20 be deleted and the paragraph start as "Appropriate interpretation of clinical test data requires consideration of contributing factors of abnormal test performance, including SES."
- P4 L51 - recommend replacing "detection of impairment" with detection of vestibular-ocular impairment"

Methods
- Could the authors provide more detail regarding the definition of Title 1 school SES. Specifically. What is high poverty and low poverty?
- P6 L4-11. This information is in Table 1 and is redundant. Please remove from text if in table. Consider adding primary sport within table or text, if available.
- P6 L48. Please delete "For the current study"

Results
- Consider adding % of athletes with abnormal NPC distance in Tables 2, 3, and 4. Authors should also consider an analyses to determine significance of those findings.

Discussion
- Authors discuss the findings of slower KD among FRL status as a result of poorer reading ability. Authors should consider expounding upon this section by discussing a higher likelihood of learning disorders among disadvantaged children. After all, KD was originally developed to assess dyslexia.
- Authors should expound upon the clinical utility of these findings. Explain to the readers why these findings matter.

Reviewer #3: The authors describe a gap in current literature that doesn't account for race and socioeconomic status in commonly used concussion assessments. Therefore, proposing a study investigating racial and SES differences on common clinical concussion assessments at baseline. Specifically, the King-Devick and VOMS assessments. The authors report no differences racial or FRL differences in VOMS pre-test or provocation scores, aside from racial differences in NPC distance and FRL status differences in VMS provocation. Similarly, no differences in K-D test were noted for any independent variable aside from FRL status, with a small effect. However, there are statistical considerations (multiple comparisons) and methodologic concerns (learning disability, dyslexia) that the authors should review. Please see the specific comments below.
Specific comments:
Abstract:
Lines 31-33: The methods does not include NPC distance as an outcome variable, which is included as one of the study aims.
Results:
Lines 37-38: Add symptom provocation to read "worse symptom provocation for visual motion sensitivity".
Line 40: Use "faster" or "slower" K-D times to describe "worse"
Conclusions:
Line 45: VMS is abbreviated for the first time.
Introduction:
Lines 11-18: this is a long list that is difficulty for the reader to digest.
Lines 33-36: Please identify clinical outcomes of what? Concussion?
Line 40: What is the century later referring to, the 1990's? That would be decades later.
Lines 51-53: Or prioritized for athletics at all.
Lines 55: change from and to or to read "free or reduced price lunch."
Methods:
Lines 35-36: Missing a period following (FRL or No-FRL).
Lines 37-40: The FRL and Title I/Non-Tile I status is confusing. Were all participants at Title I schools FRL, if this is the case why not evaluate Title I status? Was this reported by the school or the participant?
Line 42: NPC distance was an outcome variable but not mentioned.
Lines 11-40: Psychometric values are missing for the VOMS, which are provided for the K-D test.
The authors account for any learning disability or dyslexia? As the K-D test is a rapid number identification and naming assessment, this could largely influence the results. This is a major limitation if it was not accounted for.
Statistical analysis:
Line 33: add test after K-D.

Line 51: The use of "KF becomes inconsistent.

Line 35-36: A series of Mann-Whitney U tests assessed differences in symptom provocation for each VOMS item, but the significance level was not adjusted which increases the chance of type I error.

Lines 40-46: Similar to the above comment, multiple one-way anovas were performed. Also, why was the two-way anova inclusive of sex and FRL and not the other independent variables? This is not clear.

What statistical tests were used to compare NPC distance between groups? Was NPC data tested for assumptions of normality?

Were assumptions tested prior to running the regressions, specifically collinearity between FRL and race. These assumptions should be addressed in the statistical analysis.

Results:

There was no mention of pre-test VOMS in the statistical analysis, this should be included in the multiple comparisons.

When using non-parametric assessments, it is appropriate to used median and interquartile range as the descriptive statistic.

Line 24: A statistical difference is noted for the VMS VOMS item, however, if multiple comparisons was adjusted for the statistical significance diminishes. Also, what is the clinical significance of a mean difference of .1 symptom provocation between FRL and non-FRL.

Table 4: VMS symptom provocation does not match the text. The authors should review tables to ensure values are represented appropriately.

Line 59: Add "or" before current use of corrective lenses.

Lines 9-18: Consider separating the long sentence stratified by females. Also, why is it that females were separated for FRL but no other independent variable?

Table 5: Effect sizes are presented for the K-D test, but no other assessment. Also, the interpretation of effect size is not provided (despite the significant finding, .3 is a small effect), and the effect sizes are not presented in the results.

Line 54: Can the authors provide rationale as to why they chose to predict FRL? What is the clinical importance of this? Why did you decide to predict NPC and not K-D test performance?

What is the rationale for the figure with K-D test and VOMS? This is not described in the text.

Figure 1: Were the outliers for FRL included in the normality assessment? How is figure 1 different from Table 5?

Discussion:

The authors should review the results, with appropriate corrections for multiple comparisons and reflected modifications in the discussion. For example, was the VMS symptom reporting really elevated?

Lines 39-42: Please describe what is meant by K-D total time and increased NPC total distance were closely associated with White individuals. What about K-D total time, worse performance or better performance? In addition, this paragraph significantly jumps around between reading ability and concussion detection.

How do the results of this study compare to other studies assessing these clinical test in a high school population, this is missing from the discussion.

Limitations:

Do you have any information on previous assessments with these clinical tests in any of the schools? Prior assessments in either group may influence the results, i.e., practice effects.

If learning disability, or dyslexia were not accounted for, this should be included as a major limitation.

Authors' response

Dear Dr. Heger, Editorial Office, and Reviewers,

Thank you for the review of our manuscript, JCTRes-D-20-00049, "Socioeconomic status and racial disparities to be considered in clinical interpretation of pre-season vestibular and oculomotor tests for concussion." We have addressed each of the reviewer's comments/revisions using track changed and have provided a response to those changes below. Thank you again and we look forward to your decision.

Reviewer #1: This manuscript proposes to investigate an important topic - racial and socioeconomic differences in performance on the VOMS and King Devick instruments for concussion. Unfortunately, the paper diverges from that aim by partly analyzing the data contrary to that aim. In particular, the logistic regression modeling for free lunch (FRL) is backwards in that FRL is not a performance measure. It should be an independent variable (as the aim states) and not the outcome. All of the modeling should involve the performance outcomes - VOMS, Near Point Convergence and K-D. The main focus should be whether race or SES is associated with performance on these measures. The models of FRL confuse the issue and take up much space in the text.

Dear Reviewer 1, Thank you for the time you took to provide critical feedback for this project. Following this review, we believe these specific comments helped to strengthen and fine tune our manuscript. Further, the authors have amended the analysis to better reflect the purpose of the study, and have modeled all analyses to reflect consistency in use of outcomes and exposures.

The authors also seem to whittle down their sample size needlessly. Since the outcomes are (or should be) modeled separately, there is no reason to exclude the 99 people who had one of the two outcome measures but not the other.

We appreciate this comment and have updated the language to be more explicit about those excluded. Within the 99 excluded, some of those individuals also included participants with selfreported ADD/ADHD and learning disability (n=63). Also, even though the outcomes were modeled separately, the authorship team wanted to keep outcomes consistent among individuals that did complete both assessments.

They make a similar exclusion based on race, eliminating an "other" group that also has 99 people. This group is not so small that it would not add to the understanding of race in the modeling.

Thank you for this comment. This was a topic that was heavily discussed and ultimately, the authorship team chose to maintain a binary race analysis. To provide accurate and culturally competent conclusions based on 'other' classification did not seem appropriate in this instance as some of the racial/ethnic groups had as low as n=7 participants and lumping several 'other' groups into a larger 'other' group is not justified at this point in time. While we agree that having an understanding of how assessments may differ between individuals of other races/ethnicities, the authorship team has elected to maintain the stance that this is necessary future research work to be done.

There also seems to be confusion between race and ethnicity (Hispanic). Latinx individuals can be either Black or White and should be included if the question was asked correctly in two parts. There are also concerns with the modeling beyond the use of FRL as the outcome. The authors continually split the sample and fit individual models by sex and race. This approach seems analogous to a fishing expedition where they are searching for additional statistically significant results.

Thank you for this comment and bringing the issue to our attention. The confusion between race and ethnicity is understood by the authorship team. Editing was done to eliminate any potential confusions within the text. Also, the authors have amended the statistical analysis to reflect more appropriate prediction models and have removed analyses where a split sample was used. Please see Page XX, Line XX for the adjusted statistical analysis section. With the adjusted analyses, the authors do not feel this is a fishing expedition given the prior interactions noted in the literature with the independent variables.

A better approach would be to examine possible interactions in the models for the whole sample in order to determine if stratification is warranted. For example, on page 14, the model for NPC distance showed a significant race effect for those who receive FRL but not those who do not. This may be a power issue not a real finding since the non=FRL group is smaller and the black non-FRL group may be quite small further reducing the power of the comparison.

Thank you for this comment and appreciate the constructive feedback. The authors have amended the analyses to better reflect the purpose of the study and better aligned modeling.

Minor concerns
There seems to be little point to including the medians in the tables since all but one is zero. A note in the text would suffice. Also, perhaps Tables 1 and 2 could be combined.

Thank you for this suggestion for improvement. To address this concern: A sentence is included at the end of the first results paragraph stating that medians and interquartile range as 0.00 for the sample and between

groups. Medians from the tables have been removed and Table 3 and 4 were combined to reflect VOMS performance. Table 2 does not fit the demographics, so it has been removed. Therefore, table 2 now reflects VOMS and NPC performance between groups.

The text under Figure 1 draws a conclusion rather than describing the figure. Conclusions should go in the text in the conclusions section.

Thank you for this suggestion. Any concluding information has been removed from Figure 1. Further, Figure 1 was updated due to amended analyses.

The "operation definitions" on Page 6 repeats much of what is in the methods on Page 5. It only needs to be there once.

Thank you for this point made. The Operational definitions on page 6 has been removed for clarity and repetitiveness.

I was surprised that almost half of the people prescribed glasses do not wear them. As someone who has worn glasses most of my life, this is perplexing as to how they function.

The reviewer brings up an excellent point and the clarification of this was originally omitted from the text. To clarify, those that did not wear their corrective lenses during the exam do not wear them during an athletic event.

Reviewer #2: This manuscript discusses race and SES differences among commonly used vestibulo-ocular measures. Findings indicated that black athletes had higher mean NPC distances and free lunch status was associated with slower K-D times and more provocation on visual motion sensitivity. The authors conclude that race and SES should be considered as modifiers across concussion baseline measures. Specific comments below:

Dear Reviewer 2 – Your review is appreciated. The constructive feedback helped us to clarify important items within the manuscript. We have worked to address individual comments below.

Abstract
Aim
The term "performance" is used too many times. Consider removing

the first reference so that it reads "racial and socioeconomic differences

on baseline…" Thank you. Revised as recommended.

Methods
Given the topic of the paper, authors could consider including the percentage of black and FRL participants in addition to the age to provide readers with context.

Excellent point made. These have been added.

Authors use the term SES without previously defining.

Thank you for bringing this to our attention. That has been updated in text.

Consider providing analyses used in a brief sentence.

Sentence was added as suggested.

Results
I would encourage the authors to use more specific terminology when describing directionality

of their findings. Specifically, "higher mean NPC distance" is more descriptive than "worse." Same for visual motion sensitivity (e.g., "higher mean provocation") and K-D times (e.g., "slower").

Thank you for this great suggestion. We agree and have included directionality terms within the results section of the abstract.

Conclusions
Authors should define terms such as VMS before abbreviating.

We agree and have added this into the abstract as well.

The point on future research does not seem to fit here and is likely better suited for the discussion section of the paper. Consider removing.

We have decided to remove the future direction statement as recommended and have added more depth to the conclusion.

Introduction
The introductory paragraph should be more specific to the paper. The reader should finish the first paragraph and have a sense of the overall topic of the paper (i.e., statement of the problem). Given that this paper does not discuss post-concussion findings, the focus on explaining concussion does not seem to fit. Consider restructuring to focus more on the increased reliance on visual and vestibular measures in concussion

management and the need to improve methodology for interpreting those measures..

I recommend that P4 L15-20 be deleted and the paragraph start as "Appropriate interpretation of clinical test data requires consideration of contributing factors of abnormal test performance, including SES."

Thank you. Revised as recommended.

P4 L51 - recommend replacing "detection of impairment" with detection of vestibular-ocular impairment

Thank you. Revised as recommended.

Methods
Could the authors provide more detail regarding the definition of Title 1 school SES. Specifically. What is high poverty and low poverty?

Thank you for asking for clarity on this item. Title I school status is a Federally defined status given to schools of which a greater majority of students in attendance quality for free-andreduced lunch. Thus, high poverty is a Title I school and low poverty is a non-Title I school. This explanation was added into the paragraph. Citations to the Federal definition were also included.

P6 L4-11. This information is in Table 1 and is redundant. Please remove from text if in table. Consider adding primary sport within table or text, if available.

Thank you. Within the text we have removed the sentences describing participant demographics and refer reader to Table 1 to reduce redundancy.

P6 L48. Please delete "For the current study"

Revised as recommended previously from deleting this section.

Results
Consider adding % of athletes with abnormal NPC distance in Tables 2, 3, and 4. Authors should also consider an analyses to determine significance of those findings.

The authors appreciate this comment and find value in it; however we chose to shy away from any abnormal VOMS scoring to maintain examination of the study aims and not sway towards abnormal performance. We would like to explore this (false-positives) as a future direction.

Discussion
Authors discuss the findings of slower KD among FRL status as a result of poorer reading ability. Authors should consider expounding upon this section by discussing a higher likelihood of learning disorders among disadvantaged children. After all, KD was originally developed to assess dyslexia.

Thank you for this important and constructive comment…and bringing to our attention a narrative that was left out of the discussion. We have added some content on this area within the discussion of the K-D.

Authors should expound upon the clinical utility of these findings. Explain to the readers why these findings matter.

Thank you for this important point. Threaded throughout the discussion and conclusion we have aimed to massage content as to why this matters. Please see conclusion directly.

Reviewer #3: The authors describe a gap in current literature that doesn't account for race and socioeconomic status in commonly used concussion assessments. Therefore, proposing a study investigating racial and SES differences on common clinical concussion assessments at baseline. Specifically, the King-Devick and VOMS assessments. The authors report no differences racial or FRL differences in VOMS pre-test or provocation scores, aside from racial differences in NPC distance and FRL status differences in VMS provocation. Similarly, no differences in K-D test were noted for any independent variable aside from FRL status, with a small effect. However, there are statistical considerations (multiple comparisons) and methodologic concerns (learning disability, dyslexia) that the authors should review. Please see the specific comments below.

Thank you for your time and effort in reviewing this important manuscript. In regards to the statistical analyses, these have been amended to better reflect the purpose/aims of the study.

Specific comments:
Abstract:
Lines 31-33: The methods does not include NPC distance as an outcome variable, which is included as one of the study aims.

Thank you. Added NPC to the methods.

Results:
Lines 37-38: Add symptom provocation to read "worse symptom provocation for visual motion sensitivity".

Thank you. Revised to read greater (worse) symptom provocation

Line 40: Use "faster" or "slower" K-D times to describe "worse"

Thank you for the recommendation. Terms of faster and slower have been utilized as recommended.

Thank you. Revised as recommended.

Conclusions:
Line 45: VMS is abbreviated for the first time.

Thank you. This has been addressed.

Introduction:

**For the 5 line item comments below – the authorship team aimed to address each item with strategic revisions of the introduction, as suggested by the minor comments. Lines 11-18: this is a long list that is difficulty for the reader to digest.

Lines 33-36: Please identify clinical outcomes of what? Concussion?

Line 40: What is the century later referring to, the 1990's? That would be decades later.

It was referring to the 1900s and this has been modified for readability.

Lines 51-53: Or prioritized for athletics at all.

Agreed.

Lines 55: change from and to or to read "free or reduced price lunch."

Changed as suggested.

Methods:
Lines 35-36: Missing a period following (FRL or No-FRL).

Thank you for pointing out this typographical error. It has been corrected.

Lines 37-40: The FRL and Title I/Non-Tile I status is confusing. Were all participants at Title I schools FRL, if this is the case why not evaluate Title I status? Was this reported by the school or the participant?

Thank you for highlighting the need to clarify these. Yes, all participants at Title I schools were FRL. FRL was just the terminology used as point of reference for data collection. We aimed to provide better clarification of the terms within the study design and explanation of variables.

Line 42: NPC distance was an outcome variable but not mentioned.

Thank you. We have added NPC distance in (a few places) and revised as recommended.

Lines 11-40: Psychometric values are missing for the VOMS, which are provided for the K-D test.

Thank you for pointing this out. Psychometric values for the VOMS have been provided for a youth sample.

The authors account for any learning disability or dyslexia? As the K-D test is a rapid number identification and naming assessment, this could largely influence the results. This is a major limitation if it was not accounted for.

The reviewer is absolutely correct and we apologize for omitting these important details in the original submission. All athletes were excluded if they had any self-reported learning disability or attention deficit hyperactivity disorder. This was added to the methods section (exclusionary criteria).

Statistical analysis:
Line 33: add test after K-D.

Thank you. Revised as recommended.

Line 51: The use of "KF becomes inconsistent.

We are assuming the reviewer meant "K-D" of which we did an overview of the entire manuscript to check for consistency of how it was written in text.

Line 35-36: A series of Mann-Whitney U tests assessed differences in symptom provocation for each VOMS item, but the significance level was not adjusted which increases the chance of type
I error.

PI Response: We have discussed and believe that we don't need to adjust the p-value given that we planned the comparison before we ran the stats to reduce the overall Type I error inflation.

Lines 40-46: Similar to the above comment, multiple one-way anovas were performed. Also, why was the two-way anova inclusive of sex and FRL and not the other independent variables? This is not clear.

Thank you for this comment. The authors have amended the analysis to better reflect the purpose of the study. This included the removal of multiple comparisons and more reliance on multivariable linear regression models to include all independent variables as suggested. Please see Page XX, Line XX.

What statistical tests were used to compare NPC distance between groups? Was NPC data tested for assumptions of normality?

The authors apologize for the omitted text outlining the statistical analysis for the VOMS NPC. NPC distance was calculated using Mann-Whitney U (univariable) and then was the outcome variable for a multivariable linear regression model with all independent variables similar to the K-D analyses. This has been amended in the text, please see Page XX.

Were assumptions tested prior to running the regressions, specifically collinearity between FRL and race. These assumptions should be addressed in the statistical analysis.

Thank you for addressing this. We agree. The assumptions of normality were addressed prior to running any statistical analyses while collinearity was assessed within the linear regression models. Please see Page XX, Line XX for the amended statistical analysis results.

Results:
There was no mention of pre-test VOMS in the statistical analysis, this should be included in the multiple comparisons.

Thank you. Revised as recommended.

When using non-parametric assessments, it is appropriate to used median and interquartile range as the descriptive statistic.

Thank you for pointing this out. A statement was added regarding medians and IQRs. For clinical interpretation, clinicians will use mean scores, therefore means were also provided.

Line 24: A statistical difference is noted for the VMS VOMS item, however, if multiple comparisons was adjusted for the statistical significance diminishes. Also, what is the clinical significance of a mean difference of .1 symptom provocation between FRL and non-FRL.

We have discussed and believe that we don't need to adjust the p-value given that we planned the comparison before we ran the stats to reduce the overall Type I error inflation..

Table 4: VMS symptom provocation does not match the text. The authors should review tables to ensure values are represented appropriately.

We have corrected this typo.

Line 59: Add "or" before current use of corrective lenses.

Revised as recommended.

Lines 9-18: Consider separating the long sentence stratified by females. Also, why is it that females were separated for FRL but no other independent variable?

Thank you for this comment. The authors have amended the analysis to better reflect the purpose of the study and thus the stratified by females section was removed completely.

Table 5: Effect sizes are presented for the K-D test, but no other assessment. Also, the interpretation of effect size is not provided (despite the significant finding, .3 is a small effect), and the effect sizes are not presented in the results.

Thank you for highlighting this important element. Effect sizes were added for the ANOVAs and multivariable linear regression analyses. Effect size for the ordinal VOMS data (with predominantly 0 scores) is consistent with previously published VOMS studies.

Line 54: Can the authors provide rationale as to why they chose to predict FRL? What is the clinical importance of this? Why did you

decide to predict NPC and not K-D test performance? What is the rationale for the figure with K-D test and VOMS? This is not described in the text.

Thank you for this comment. After careful discussion and decision to better align analyses with the purpose statement, the model to predict FRL was removed and the authors have amended the analysis with the inclusion of a multivariable linear regression analysis with the K-D [and NPC] as the outcome variables.

Figure 1: Were the outliers for FRL included in the normality assessment? How is figure 1 different from Table 5?

Thank you for this comment. Table 5 was deleted due to the amended statistical analysis and Figure 1 has been clarified and updated, including the outliers assumption.

Discussion:
The authors should review the results, with appropriate corrections for multiple comparisons and reflected modifications in the discussion. For example, was the VMS symptom reporting really elevated?

Thank you for this suggestion - citations for methodology have been included to explain analyses no corrected for multiple comparisons.

Lines 39-42: Please describe what is meant by K-D total time and increased NPC total distance were closely associated with White individuals. What about K-D total time, worse performance or better performance? In addition, this paragraph significantly jumps around between reading ability and concussion detection.

Thank you for suggesting revision for clarity in this paragraph. The language noted in the reviewer comment was significantly revised and much of it removed to provide readers with more clear thoughts.

How do the results of this study compare to other studies assessing these clinical test in a high school population, this is missing from the discussion?

Thank you for pointing out the need to compare results with previous studies in high school populations. Some content was added into the discussion.

Limitations:
Do you have any information on previous assessments with these clinical tests in any of the schools? Prior assessments in either group may influence the results, i.e., practice effects.

We appreciate this comment and understand the value of knowing any potential for practice effects. This baseline testing program was in fact the first time any of the high school athletes had formal baseline testing within their high school framework. So there were no formal prior assessments or influences of practice effects.

If learning disability, or dyslexia were not accounted for, this should be included as a major limitation.

Both LD and dyslexia were accounted for and any participant that reported having one of these were excluded from analyses.

2nd editorial decision
29-Aug-2020

Ref.: Ms. No. JCTRes-D-20-00049R1
Socioeconomic Status and Racial Disparities to be Considered in Clinical Interpretation of Pre-Season Vestibular and Oculomotor Tests for Concussion
Journal of Clinical and Translational Research

Dear Dr. Wallace,

Reviewers have now commented on your paper. You will see that they are advising that you revise your manuscript. If you are prepared to undertake the work required, I would be pleased to reconsider my decision.

For your guidance, reviewers' comments are appended below.

If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript.Also, please ensure that the track changes function is switched on when implementing the revisions. This enables the reviewers to rapidly verify all changes made.

Your revision is due by Sep 28, 2020.

To submit a revision, go to https://www.editorialmanager.com/jctres/ and log in as an Author. You will see a menu item call Submission Needing Revision. You will find your submission record there.

Yours sincerely

Michal Heger
Editor-in-Chief
Journal of Clinical and Translational Research

Reviewers' comments:

Reviewer #3: The authors did a good job in making revisions for this manuscript. I believe that this manuscript should be published, with the suggested analytical revisions as this adds to the current literature that is lacking in respect to diverse sociodemographic groups. With this revision, the authors would add to the literature that there are not differences in the VOMS assessment in healthy athletes within this range, suggesting more work done in injured populations. And if differences do exist after injury, the lack of baseline difference would foster further research with in depth examination for the "why."

Introduction:

The authors made appropriate revisions to the introduction.

Methods and Results:

The authors clarified most of my method concerns.

I do however, disagree with the authors failing to control for multiple comparisons. Although they set the critical α to .05 a priori, to maintain this overall probability of making a type I error for each comparison, the α should be divided by 8 VOMS test components assessing symptoms (pre-test, smooth pursuits, horizontal saccades, vertical saccades, convergence, horizontal VOR, vertical VOR, VMS). This would result in an adjusted critical α at .006 for each comparison. Currently, the authors are setting the α for each test at .05 so they have inflated the probability of making a type I error to an overall $\alpha = .40$. However, it is apparent that as the number of comparisons increases, the adjusted critical α becomes increasingly small, so the chances of finding a significant difference becomes miniscule. Therefore, the Bonferroni correction is not universally accepted, and in any case, is considered to be untenable when more than 5 comparisons are being made. Thus, at minimum the authors should set the critical value at $p < .01$ for each comparison. In doing so, the VMS score should be interpreted with $p < .01$.

Hazra A, Gogtay N. Biostatistics Series Module 3: Comparing Groups: Numerical Variables. Indian J Dermatol. 2016;61(3):251-260. doi:10.4103/0019-5154.182416

The correction for multiple comparisons has been done correctly in previous work.

Elbin RJ, Sufrinko A, Anderson MN, et al. Prospective Changes in Vestibular and Ocular Motor Impairment After Concussion. J Neurol Phys Ther. 2018;42(3):142-148. doi:10.1097/NPT.0000000000000230

Discussion:

The authors do a nice job of interpreting the clinical meaningfulness of the VMS differences that were reported to be significant. However, should consider revising the results, and therefore making adjustments to the VMS interpretations to the discussion.

Tables and Figures:

There is still a figure included that is not described in the text on page 12, and does not have a title or description. From a viewer perspective, this black figure illustrates that the K-D test does not equal the VOMS test. The clinical meaningfulness of the VMS differences reported in the results also may not be appropriate as this figure stand alone does not indicate the very small differences reported. I suggest it is removed or revised.

Authors' response

Dear Dr. Heger and Reviewers:

Thank you again for the opportunity to further revise this manuscript and provide justification for analysis decisions. Please see our responses to your comments and suggestions below in

*red font. The authorship thanks each individual reviewer to helping to
strengthen this work to add to the clinical utility of concussion measures.*

Reviewer #3: The authors did a good job in making revisions for this manuscript. I believe
that this manuscript should be published, with the suggested analytical revisions as this adds
to the current literature that is lacking in respect to diverse sociodemographic groups. With
this revision, the authors would add to the literature that there are not differences in the
VOMS assessment in healthy athletes within this range, suggesting more work done in injured
populations. And if differences do exist after injury, the lack of baseline difference would
foster further research with in depth examination for the "why."
Introduction:
The authors made appropriate revisions to the introduction.

Thank you.

Methods and Results:

The authors clarified most of my method concerns.
I do however, disagree with the authors failing to control for multiple comparisons. Although
they set the critical α to .05 a priori, to maintain this overall probability of making a type I
error for each comparison, the α should be divided by 8 VOMS test components assessing
symptoms (pre-test, smooth pursuits, horizontal saccades, vertical saccades, convergence,
horizontal VOR, vertical VOR, VMS). This would result in an adjusted critical α at .006 for
each comparison. Currently, the authors are setting the α for each test at .05 so they have
inflated the probability of making a type I error to an overall $\alpha = .40$. However, it is apparent
that as the number of comparisons increases, the adjusted critical α becomes increasingly
small, so the chances of finding a significant difference becomes miniscule. Therefore, the
Bonferroni correction is not universally accepted, and in any case, is considered to be
untenable when more than 5 comparisons are being made. Thus, at minimum the authors
should set the critical value at $p < .01$ for each comparison. In doing so, the VMS score
should be interpreted with $p < .01$.

*After careful discussion and consultation with a statistician, we have decided to add a test for
multiple comparisons in the use of the Benjamini-Hochberg procedure for false-discovery
rate. We have included these adjustments within the language in our statistical analyses,
results, and discussion sections.*

Hazra A, Gogtay N. Biostatistics Series Module 3: Comparing Groups: Numerical Variables.
Indian J Dermatol. 2016;61(3):251-260. doi:10.4103/0019-5154.182416
The correction for multiple comparisons has been done correctly in previous work.
Elbin RJ, Sufrinko A, Anderson MN, et al. Prospective Changes in Vestibular and Ocular
Motor Impairment After Concussion. J Neurol Phys Ther. 2018;42(3):142-148.

doi:10.1097/NPT.0000000000000230

Discussion:

The authors do a nice job of interpreting the clinical meaningfulness of the VMS differences that were reported to be significant. However, should consider revising the results, and therefore making adjustments to the VMS interpretations to the discussion.

Thank you again. We have revised the discussion with updated results of adjusted analyses. The VMS is no longer significant and thus we have cut down on the discussion of this finding within the discussion. The discussion as a whole has been trimmed.

Tables and Figures:

There is still a figure included that is not described in the text on page 12, and does not have a title or description. From a viewer perspective, this black figure illustrates that the K-D test does not equal the VOMS test. The clinical meaningfulness of the VMS differences reported in the results also may not be appropriate as this figure stand alone does not indicate the very small differences reported. I suggest it is removed or revised.

The figure was the graphical abstract. The graphical abstract has also been updated and placed directly after the written text abstract.

Statistical review of

JCTRes-D-20-00049R1

Results:

Page 9, line 11. "A series of Mann Whitney U tests revealed no differences". This is false. The P value may have been high, but the two groups were not identical so it is not true that there are no differences. Again on line 18 "no differences". The difference was small and consistent with random sampling, is what you mean to say, I think.

This language was adjusted in text within the results section as suggested by the above interpretation.

Figure 1 shows mean and SD. Much better to show some detail. A scatter plot of every point might be too busy to be understandable, but a violin plot would be helpful. A box-and-whisker plot doesn't show as much detail as a violin plot, but a lot more than mean and SD.

You should also show the comparison of NPC scores in the same form as
new Figure 1 (violin plot).

*The original Figure 1 was removed. Violin plots have been created as suggested, for both K-
D times and NPC scores. Thank you for this excellent suggestion to disseminate results.*

"None of the variables in the model had any collinearity (no relationship) to one another. "
This can't be true. I think you mean the collinearity is small and consistent with random
sampling. But I find it impossible to believe there is no association (now in the US) between
race and SES.

*This language has been modified to better align with updated statistical analyses. Assessment
of multicollinearity was assessed using VIF and that process is now included within the
results section.*

Table 3. The coefficient for race is -2.14 seconds for KD time. Figure 1 shows that the mean
KD time is about 50. So this is a 4% difference. Is this enough to matter? Hard to imagine.

*Table 1 has been updated due to the revised analysis plan. Further, Figure 1 was removed
and violin plots for this data have now been included as suggested by the reviewer.*

You state that the model only explains 3% of the variation. I think it would be good to
highlight this with a graph (that ought to be shown with every multiple regression, in my
opinion) of actual KD time (X) vs. KD time predicted by the model (Y). If the model
explained a lot of variation, the points would all hug a 45 degree line. If the model explains no
variation, the points will show no pattern, which is what I think these data will show. (And the
same for the other multiple regression.)

*Thank you for this point made. We did not represent this using a figure or graph; however,
this analysis has been revised significantly with the help of a biostatistician.*

Both multiple regressions assume the residuals are Gaussian. Please address this assumption.
Especially for the NPC data which you say are not normal. Consider showing residual plots.

Page 9: "All medians and interquartile ranges were 0.00 for VOMS symptoms and NPC

distance in the sample and between groups." Huh? I don't understand what you mean or how this can be. You say data are not normal, but don't show the distribution. This comment really makes me want to see the actual distributions you are comparing.

Due to the nature of the concussion tool, this is common of the VOMS [medians and interquartile ranges] to see ranges of 0.00.

Terminology

You use the word "significant" 22 times. This term is widely misunderstood. I'd suggest avoiding it and discussing the actual results. Fine to show P values, and talk about rejected null hypotheses when that makes sense. A few "statistically significant" mentions wouldn't be so bad. But people tend to think that "significant" means important or large or having consequences, when it really just means one number (p) is less than another (preset alpha). Discuss the findings not the p values.

Thank you for pointing this out. We have worked to reduce the use of significant per this comment.

Graphical abstract.

The resolution is too low so I can't read most of it. Maybe just a document conversion issue, but I think the print is simply too small to be readable (and yes, I wear my prescribed glasses!).

This has been completely revised. The new graphical abstract can be seen directly after the written abstract.

Overall:

I didn't read the entire discussion and don't know this field.

It seems to me that you have shown that the association of race and KD time is, at most, tiny. Looked at by itself, you get a high p-value. Looked at in a multiple regression with other variables, you tease out a small p-value but an average difference of only about 4% of the mean KD time. Isn't this your main finding, that any difference is tiny?

We have updated the language for the updated regression model – and have included discussion of this finding within the text of the discussion section.

For NPC the difference is greater (about 30%; last row of table 2), but the p value is not all that small (0.02) and it is one of ten comparisons, so when corrected with Bonferroni, it is far from convincing. You don't show a graph of the NPC distributions in the two groups, but should. Especially since you say it is far from Normal and that the median is zero.

A violin plot for NPC has now been included. Also, adjusted values are now included following Benjamini-Hochberg procedure for false-discovery rate.

Multiple comparisons

I was asked to comment about multiple comparisons…

Multiple comparisons. Essential questions to interpret the data:

-How many statistical comparisons did you do (how many p-values)?? About 30, I think.

-Were these all planned?

-Were all planned comparisons published?

-Did the study design designate one (or a few) comparisons as the key (primary) outcome, and others as secondary?

These were all planned at the start of the study. Adjustments were made using the Benjamini-Hochberg procedure for false discovery rate.

With 30 p-values, you expect some small ones by chance. If you did some p-hacking with the data, and so reported unplanned comparisons or only reported the run of the multiple regression that gave you the results you wanted, then the results really cannot be interpreted.

No p-hacking was done; and unplanned comparisons were not completed.

When to apply Bonferroni (or other) correction is tricky. I think it is clear it should be used among a set of related comparisons. But which comparisons are related? The convention in many fields is all those shown in one figure or table, but that is a bit arbitrary. Some people have said that corrections are not needed when comparisons were planned, but that doesn't make logical sense to me when a large set of comparisons were planned (as was the case

here). Multiple regression nearly always reports a bunch of p-values, but these are never corrected for multiple comparisons. Just tradition. Ditto for two (or more) way ANOVA. So there are no clear rules, except to report clearly what was done.

This paper, I think, should emphasize the size (small) of the effects, and deemphasize p-values. It seems to me that the main finding is that race and ses make very little difference, but the differences seen may be more than random sampling would provide.

Thank you for this important comment. Discussion of effect-sizes was incorporated more within the results (Cohen's d were also computed) and within the interpretation and clinical utility.

When summarizing effect sizes, I prefer to show the effect in the actual unit of the measurement or the percent change (when the variable is a ratio variable). Normalized effect sizes (Cohen d) show the ratio of the difference divided by the SD. While this is helpful for combining (meta analysis) different kinds of data, I don't think it helps a reader understand the data in this paper as much as reporting percent change.

Thank you for this comment and clarification. Percent change would be a very important measure post-concussion. This data reflects healthy baseline values. Cohen's d effect sizes were computed and included within the results, and interpretation of results for this reason.

Regarding multiple comparisons (the reason I was asked to comment), how about:

- Clear statement about a clear statement of how many p-values were computed, how many were planned, how many planned p values were reported, and the degree to which the analysis protocol was planned vs. evolved as the data were looked at.
- State that with 30 comparisons, you expect (30 * 0.05) 1 or 2 to be <0.05 just by chance (if they are independent). You have more small p values than that so unlikely to be all a matter of multiple comparisons.

All analyses were pre-planned and statements now reflect that to minimize confusion. Also, language included about the adjustment using Benjamini-Hochberg procedure false-discovery rate were included. Due to these corrections, the discussion was trimmed and revised accordingly.

3rd Editorial decision
24-Sep-2020

Ref.: Ms. No. JCTRes-D-20-00049R2
Socioeconomic Status and Race as Social Determinants of Health to be
Considered in Clinical Use of Pre-Season Vestibular and Oculomotor Tests for Concussion
Journal of Clinical and Translational Research

Dear authors,

I am pleased to inform you that your manuscript has been accepted for publication in the
Journal of Clinical and Translational Research.

You will receive the proofs of your article shortly, which we kindly ask you to thoroughly
review for any errors.

Thank you for submitting your work to JCTR.

Kindest regards,

Michal Heger
Editor-in-Chief
Journal of Clinical and Translational Research