Journal of Clinical and Translational Research
Peer review process file 07.202101.013

Reliability and concurrent validity of TRAZER compared to three-dimensional motion capture

Jennifer A. Hogg, Lynette M. Carlson, Abigail J. Rogers, Mason W. Briles, Shellie N. Acocello, Gary B. Wilkerson

Corresponding author
Jennifer A. Hogg
Department of Health and Human Performance, The University of Tennessee Chattanooga, Chattanooga, TN, USA

Handling editor
Michal Heger
Department of Pharmaceutics, Utrecht University, the Netherlands
Department of Pharmaceutics, Jiaxing University Medical College, Zhejiang, China

Review timeline:

Received: 9 September, 2020
Editorial decision: 14 October, 2020
Revision received: 13 November, 2020
Editorial decision: 6 December, 2020
Revision received 12 December, 2020
Editorial decision: 13 December, 2020
Published online: 25 February, 2021

1st Editorial decision
14-Oct-2020

Ref.: Ms. No. JCTRes-D-20-00088
Reliability and Concurrent Validity of TRAZER Compared to 3-Dimensional Motion Capture
Journal of Clinical and Translational Research

Dear Dr. Hogg,

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.

Please pay particular attention to the fact that a number of technical details of both systems were ignored or missed. Such details should explain much of the results, particularly with respect to the error measures between the systems, although the intraclass correlation values were high in two measures and with no correlation in one measure. The number of variables to compare both systems are limited to draw conclusions. Also, the way the study was carried out does not test the reliability of the system.

Further note that some reviewer comments are attached as a separate document.

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 Nov 13, 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: Dear authors -

Please see my direct comments on the manuscript in the attached document.

Reviewer #2: Thank you for asking me to review this manuscript. Herein, this study investigated the reliability and concurrent validity of the TRAZER system compared to Vicon 3D motion capture system during the TRAZER task battery in healthy young adults. Overall, the findings and scientific value of this study is appropriate for the Journal of Clinical and Translational Research. The manuscript is lacking expansion on protocols and concepts, however, with edits the concepts can be well-linked to the value of these data to clinicians and research scientists. Please see line specific suggestions and comments below.

Abstract
- Lines 7-9: First sentence is a bit of a run-on sentence; consider breaking into two sentences and linking them with a connecting phrase. (Ex: injury… As such, clinically…)
- Line 12: Consider addition of space requirement and/or software licenses to cons against motion capture
- Authors mention collecting maximum speed, but don't mention this in the methods (mention maximum velocity). As mentioned below under specific concerns for the Methods, please clarify which of these variables were collected as they mean different things.
- No key words listed
Introduction
- Lines 2-3: Consider expanding on what the functional movement assessment typically consists of (e.g., gait analysis)
- Lines 2-6: Citation needed
- Lines 7-9: This statement could have increase weight if the authors provide statistics on days lost to injury, health care costs, long term injury risk (e.g., knee OA)
- Line 10: The first sentence can be omitted

- Line 13: Citation needed
- Lines 16-20: Feed forward mechanisms are important as well
- Line 18: I like this sentence as it makes a great point, however, it can be expanded upon. For instance, unanticipated events can lead to improper muscle tone and joint stiffness, leading to a sensory prediction error, and thus a musculoskeletal injury.
- Line 31: citation needed
- Line 40: Avoid colloquial phrases like "on the other hand"; consider "conversely"
- The introduction should include more information on TRAZER (i.e., what does it measure, how, etc.)

Methods
- Line 51: Please expand on "time and space constraints" which limited data collection
- Line 76: Operationally define the base of the spine
- Line 78: Citation
- The authors do not justify their test-retest interval of 3 consecutive days of testing, as typically, there is a longer time interval between each assessment time. Some expansion on this might benefit the reader as a concern arises for a potential learning effect (although the results disagree with that; p=0.32).
- Maximum speed is not mentioned a variable collected in the methods section (but maximum velocity is). Please clarify which of these variables were collected via TRAZER and 3D motion capture.
- Statement needed on IRB approval in addition to existing informed consent statement
- A subsection called "Instrumentation" would be beneficial to the reader whereby the authors can further explain the TRAZER system. For example, how does it capture?

Discussion
- Lines 118-123: These sentences can be more concise, all the while stating the same thing.
- Lines 123-125: Please explain the translational and clinical relevance of this finding (as to align with the journal's mission)
- Line 131-132: Should be Lopes's since it's a singular possessive noun
- Lines 134-136: This is truly the take home message of this finding, as such, the reader would benefit from an expansion on these ideas and sentences.
- Line 138: Define more affordable (how much more?)
- Line 141: "Nyman, n.d.": No date? This is a M.S. Thesis project and the comparison isn't appropriate for a full paragraph relative to a conference abstract.
- Line 155-156: As mentioned in the methods, the fact that Vicon was sampled at 60 Hz seems like a poor reasoning in error, as the sampling rate could have easily been
- The concept of TRAZER's proprietary algorithms should be expanded upon as I view this as a large deterrent from using TRAZER. Can the user trust this "black box" method? Additionally, the differences in methods/results between the study herein and Nyman's should be expanded upon further.
- The difference in sample rates could have been avoided, why did the authors not change the Vicon system sampling rate to 30 Hz for ease of comparison? This seems like a study design oversight that warrants further discussion.
- Were participants tested at the same time each day? If not, this could be added to the limitations as reaction time may differ at 8AM vs. 3PM.
- Again, in reference to the journal's mission, the discussion needs a clear clinical and/or translational aspect to the study

References:
- Line 226: Can the Nyman source be updated further? Is this a website, book, journal, op-ed?
- We there any power analysis conducted? (https://pubmed.ncbi.nlm.nih.gov/24197712/)

Tables & Figures:

- Figure 2: Crop graph to remove the bar on the left outer edge of the graph
- As mentioned above under Abstract and Methods, there seems to be a confusion on variable name (maximum speed vs. maximum velocity); please clarify which of these were used in Table 2 and Figure 3.

Reviewer #3: This sort of validation paper has clinical utility, especially given the need for low cost portable systems to monitor rehab progress. There is critical detail missing from the background and methods section, however, which requires major modification prior to publication. Specific comments are below:

Background:

1. Some additional background on the TRAZER system and its mechanism of kinematic data capture is needed. Kinect is mentioned here. How does TRAZER compare to the Kinect? The gap in research can't just be that this is a new system so we should test it also. There are a lot of new systems, why not one of them? TRAZER is chosen why? It is novel how?
2. Prior work on the Kinect was appropriately mentioned in the background. It may be helpful to state why Kinect has "variable test-retest reliability." Was the Kinect's variable reliability related to which kinematic variables were measured? If reliability of measurement may be a limitation of the kinect in some settings or for some applications, as is implied here, what are the potential advantages of TRAZER that may make it theoretically more reliable than the Kinect (if any)?

Method:
1. "Due to time and space constraints". Be more specific, what time and space constraints? Otherwise, omit.
2. "participants were not digitized" What does that mean?
3. It is unclear from the write-up what data points the TRAZER system provided. Raw coordinates of the S2 joint only?
4. If TRAZER calculations are proprietary, how can it be assured that the same calculations are being used by the TRAZER system and for the Vicon data?
5. Did the researchers calculate the metrics for both datasets from available outputs of the TRAZER system?
6. The clinical relevance of each of the outcomes of interest should be explained, e.g., why would total distance be of interest to a clinician? E.g., joint angles such as knee flexion during a squat have more obvious relevance than the speed and location of a particular body joint. The clinical rationale needs to be clearer in both the Intro and the Methods section.

Results:
1. The Bland Altman plots seem to show a proportional bias, i.e. that the errors are proportional so that the larger values have more error compared to Vicon. It is presumed that the measurements of the TRAZER system are on the x-axis, but this should be stated.

Discussion: The utility of the TRAZER system should be contrasted with other available systems for which reliability and validity have been measured, e.g., Kinect.

Reviewer #4: The manuscript brings a relevant topic in human movement research. However, there are a number of issues that compromises its quality.
The introduction must be assertive towards the Trazer System. The first two paragraph does not add much to the rationale of the study. The authors could bring technical details of the system and studies done with it, for instance, to demonstrate the necessity of the present study. The third paragraph is too long. It can be broken in two and the authors should concentrate on it to improve the rationale of the study. For instance, the explanation of how the Trazer System tracks and reconstruct movements within the cartesian plane would be relevant. A number of processes of the way the cameras of the Trazer system capture body's motion and the algorithm reconstructs it in bi-dimensional or tri-dimensional perspectives are relevant to understand the comparisons with Vicon system. With respect to the rationale of the study, it is reasonable to bring details of each system, so we will be aware of the differences of each system. To better suit a wide range of interested reader I suggest to explain the theoretical meaning of reliability and validity. With respect to the reliability testing is not clear why that is relevant? The manner in which reliability testing was designed in the present study is not a straightforward measure of it. Actually, the protocol tested differences in the measures of performing the task with the Trazer system in three different days. The validity testing is the relevant to test the Trazer system with an accurate system.

The procedures of data collection need more details and better organized. Separate the explanations of the two experiments in different paragraphs or topics. Provide details of the procedures for both experiments (reliability and validity).
Operational definition of reliability and validity should be included in procedures. The sentence on line 59 "to obtain test-retest reliability data, participants were not digitized." does not make sense without a proper explanation. How many trials or repetitions did the participants perform in each day for the first experiment (i.e., reliability)? Was the order of repetition random? Was there any procedure to calibrate each participant's body? Explain the technical aspects of data collection with the Trazer system. How does the system recognize a human body and its movement per segment? Does the Trazer system reconstruct the movement in bi-dimensional or tri-dimensional perspective?
Start the description for the validity procedures in a new paragraph to separate from the reliability experiment. Line 71, the sentence "…they were digitized" needs a concise and clear explanation. Explain whether the Vicon system and Trazer system are tracking the exact same point on the body? And how are they tracking the points according to their cameras? What is the accuracy of each camera's system (Trazer and Vicon)? Although the data reduction from Trazer System is built in and apparently there is no possibility the authors treat the raw data, then the authors must describe the process of treatment of raw data by the built-in algorithm.

Discussion section needs to improve in a way that the authors must demonstrate differences that each system may produce as a result of data treatment. On line 155 the authors may suspect about the differences in sampling rate between both systems. That is probably the reason of differences between the systems. However, a detailed of data treatment and the algorithm used by each system to calculate each variable would provide a clue of the lack of correlation. For instance, how each system calculated each variable? If Trazer has one camera and Vicon has 8 cameras. The discussion is confusing. For example, '…high absolute agreement…' (line 137), that is not adequate to say when the ICC was carried out and not Cohen's Kappa. Line 155: what the authors want to say with lack of congruity? Were the results about congruity presented in the results section? I did not agree that motivation of the participants would be a limitation of the study. Actually, the limitation is on the procedure or technique used by the authors to examine the reliability of the system.

There is additional documentation related to this decision letter. To access the file(s), please click the link below. You may also login to the system and click the 'View Attachments' link in the Action column.

Authors' response

November 13, 2020

Reviewer #1,

Thank you for your thorough and thoughtful comments on our manuscript. We appreciate the opportunity to address the identified shortcomings, and feel that our manuscript is much stronger as a result. Our responses to each of your comments are detailed below. Additionally, changes within the manuscript have been highlighted for ease of location.

P4, L12: the term "non-immersive virtual reality" should be "augmented reality" per the industry standards. Please adjust.

Response: Thank you for this comment. Your comment informed us to provide further details on the system and camera in lines 32-39 of the introduction. The TRAZER system does not augment or enhance objects in the real world – our understanding of augmented reality. Prior publications (Nyman 2017, Wilkerson et al. 2018) recognize TRAZER as being non-immersive (without goggles) virtual reality. The TRAZER camera does recognize a person standing in the field of view and represents that person with a simulated model (avatar).

P4, L15: independent evidence of the validity and reliability may come off as biased by the research team to prove the TRAZER is inaccurate. Thus, I would suggest removing this sentence and stating that test-retest and concurrent validity are needed.

Response: Revised as recommended. Sentence in abstract changed to "Test-retest reliability and concurrent validity of these systems is currently lacking."

P5, L18-23: Since the system being used for assessment is augmented reality, the use of term "virtual reality" is not appropriate and should be changed. Furthermore, a possible comparison or information on augmented reality (i.e. a system like the wii fit or the Bertec half-dome system) would be more appropriate for the introduction. Please revise.

Response: Please see comment above regarding virtual reality. TRAZER utilizes a Kinect camera, which has been studied extensively. With the help of review comments, we focused on expanding the literature review of the Kinect camera in lines 44-50. Wii Fit and Bertec utilize external devices (balance board and treadmill), while the TRAZER only requires interacting with visual markers displayed on the TV monitor. We revised by expanding on the description of the system.

P6, L29: the term non-immersive is inaccurate as it is augmented reality. I think the information on non-immersive is helpful but it could be enhanced by using the correct term.

Response: Edited as above. We also expanded on the literature review of the Kinect studies.

P7, L34-38: the jump from postural control to kinematics makes the introduction a bit jumpy. Since you are really only comparing direct marker movement, I would prefer to stay in the realm of kinematics unless the postural data is relevant. If not, delete and briefly discuss the efficacy of motion tracking data.

Response: Thank you for this comment. We believe the postural data is relevant because it does utilize the Kinect camera and assesses kinematic angles, albeit during a postural task. For reliability and validity purposes, we consider this appropriate. Revised line 44 to explicitly state the connection between the Kinect and the TRAZER.

P7, L42: as stated in the abstract comments, remove the phrase independent evidence. It comes off a bit bias. Just state the objective and that no research has been conducted in the manner previously. In addition, circling to the discussion the Nyman study (with no date?) should be discussed in the introduction as the only supporting evidence and that this data should be expanded to include a more robust dataset etc.

Response: Revised as recommended and consistent with the abstract. Nyman study expanded upon, including the date. Revised lines 60 and 66.

P7, L53-54: please provide sex information along with the anthropometrics.

Response: Revised as recommended. Revised lines 82-83: "The reliability cohort consisted of 18 healthy individuals (7 male, 11 female…) The validity cohort consisted of 13 healthy individuals (5 male, 8 female…)"

P7, L55-57: was prior concussion history included in the injury history? That is an important distinction, as the outcomes of neuromechanical control will be confounded. It should not directly influence the overall study results but it important to report.

Response: Revised as recommended. Thank you for this comment. While we agree that concussive history doesn't directly influence the aim of the study, it would caution the reader against extrapolating our data as normative. Revised line 87: "History of concussion was not considered exclusionary." This was also included as a limitation on lines 256-258: "History of concussion was not considered exclusionary for participation. While this does not affect test-retest reliability or concurrent validity, the reader should caution against using these data as reflective of the larger population."

P7, L60: why was 3-day test-retest reliability performed rather than the traditional 1-to 2-week assessment periods? Assessments over a short period could lead to a learning effect across the data. Please justify.

Response: Revised as recommended. Revised lines 111-112: "Relying on pilot testing and anecdotal accounts, three consecutive days were used to account for a possible learning effect on the first day."

P7, L65: Figure 1 is not entirely descriptive of the process of how the TRAZER works. Could a series of shots demonstrating the movements to a few of the eight possible locations (or a graphical sketch?) would be very helpful. In addition, was the final position specified across conditions? For example, what was the total displacement for each location and did it vary across conditions.

Response: Revised as recommended. A new Figure 1 was created according to your suggestion. Total displacement of each position was constant across conditions and participants, with 40 total repetitions. Thus, each participant had the same 'traveling' requirements. This was made more clear in revised lines 90-104.

P8, 72-73: if the S2 spinous process was the most important marker for the TRAZER, why include the others? Please justify. In addition, since you are indicating bony landmarks please give the correct anatomical terms for the locations.

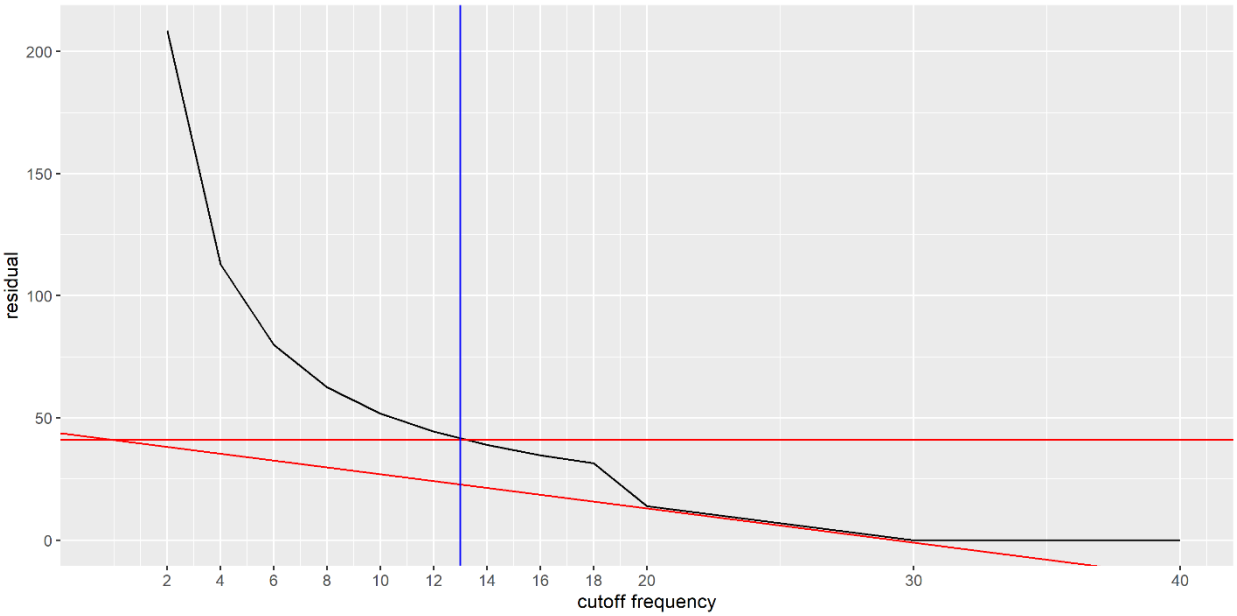
Response: Markers were placed on either foot to conduct secondary analyses, which never materialized. This information was deleted. Revised lines 114: "…in which a reflective marker placed over the S2 spinous process..."

P8, L74-75: why are the Hz set at varying conditions? Pending the Vicon camera, it can easily collect at 30Hz. This could explain some of the large variations in the concurrent validity data and could increase accuracy of the motion capture data compared to the TRAZER. Please justify.

Response: We do agree that the difference in sampling rates likely accounts for some of the accuracy difference between systems. However, as a gold standard, 3D motion capture systems typically sample at higher frequencies, thus this was the comparison we chose to make. Given that the length of each trial was ~3 minutes, 60Hz was the highest frequency we felt comfortable would not overwhelm the system. Nevertheless, we do concede that this is a limitation and warrants further discussion. We have added this as a limitation in revised lines 249-253: "An important limitation of this study was the different sampling rates of each system; TRAZER captured at 30 Hz, while Vicon captured at 60 Hz. We acknowledge that this difference may partially account for discrepancies in accuracy. However, as gold standard 3-dimensional motion capture typically samples at higher frequencies, this is a more externally valid comparison." and expanded the discussion in revised lines 230-237: "The authors elected for a 60 Hz Vicon sampling rate because this is more representative of a gold-standard motion capture collection. This is consistent with other researchers who have used different sampling frequencies when comparing two-dimensional with three-dimensional motion capture (Clark et al., 2012; Maykut et al., 2015; Nyman, 2017). In fact, in the white paper reported by Nyman, three dimensional motion capture was sampled at 120 Hz, while TRAZER was sampled at 30 Hz. Thus, although different sampling rates likely partially explain the systematic differences observed between the TRAZER and Vicon in the current study, it does not fully account for the discrepancy."

P8, L82: why us a 12Hz low-pass filter? Why not a 10Hz or 18Hz filter? Can you provide the FFT of the data to support the use of the filter? Please justify.

Response: Winter's 2nd edition text recommends a residual analysis to determine cutoff frequency, the results of which are below. You can see a jump in the noise : signal ratio from 18 to 20 Hz. To err on the conservative side, we drew the tangent line on the lower end. This led to an optimum frequency of 13 Hz (blue line), which is close to the 12 Hz frequency we used. The following text was added in lines 126-127: "The cutoff of 12 Hz was chosen after a residual analysis indicated 13 Hz to be the optimum cut point to maximize the signal to noise ratio."



P8, L86: Unless I misread this information, the maximum data is the single best trial? Is this correct? I always prefer equations within the text if possible and providing that would be ideal over stated computations.

Response: Revised as recommended. The following equations were substituted in lines 132-137:

Total distance (m) = $\Sigma\,|p_i - p_{i-1}|$, where p = position and i = capture frame

Maximum velocity (m/s) = $\frac{\max(p_i - p_{i-1})}{t}$, where p = position, i = capture frame, and t = time

Maximum acceleration (m/s^2) = $\frac{\max(v_i - v_{i-1})}{t}$, where v = velocity, i = capture frame, and t = time

P9, L96: the use of paired t-tests are not preferred over an omnibus test. Please re-run using an appropriate ANOVA model (RM ANOVA possibly) to ascertain f-values.

Response: Revised as recommended. Wilks' Lambda was derived from a one-way multivariate repeated measures analysis of variance (MANOVA) to perform an omnibus test, with P<.05 used as the standard for a difference among the 3 trials. Follow-up testing involved separate univariable repeated measures analysis of variance (RM-ANOVA) for each of the 8 dependent variables. Revised lines 155-159 and 161-163: "A repeated measures MANOVA including all eight dependent variables demonstrated a significant difference among trials (Wilks' Lambda = 0.36; $F_{15,54} = 2.23$; p = .015)…Test-retest consistency among trials, Shapiro-Wilk test of normality result for each trial, and unvariable repeated measure analysis of variance results for differences among trials are presented in Table 1." For validity, revised lines 167-169: "A repeated measures MANOVA including all three dependent variables demonstrated a significant difference between systems (Wilks' Lambda = 0.03; $F_{3,10} = 94.00$; p < .001)."

P9, 103: was the data assessed for normality and influential skewness? Data in Figure 2 and 3 indicate that you had possible outliers in your dataset. Thus, please report the normality of the data. If normality is not assumed, use the appropriate non-parametric statistics.

Response: The Shapiro-Wilk test was used to assess distribution normality and the results have been included in Tables 1 and 2.
Because neither logarithmic nor square root transformation provided any substantial improvement in the Vicon distributions for total distance and maximum velocity, the results of non-parametric correlation (Spearman's rho) were included in Table 2.
Regarding reliability, lines 155-157 were revised: "A significant deviation from normality was evident for day one's reaction time (Shapiro-Wilk p=.002), but neither logarithmic nor square root transformation provided any substantial improvement in distribution normality." Revised lines 164-167, regarding validity: "Vicon distributions for total distance and maximum velocity deviated significantly from normality, but neither logarithmic nor square root transformation provided any substantial improvements. Thus, the results of non-parametric correlations (Spearman's rho) were also calculated and are presented in Table 2."

P9, L110: please report the raw data by each day along with the SEM in Table 1 please. This can help with future studies that may want to replicate your results.

Response: Revised as recommended. Table 1 now includes descriptive data for each of the three days.

Table 2: it would seem that for each trial the participants travel a total of 7 to 8 feet (40 trials in a 25 sq ft space). Is this correct? If so, I think this is vital information that can be added to the methods section to describe the total movement during the 40 trials. It may also help to support your effort conclusions.

Response: Yes, this is correct. Revised as recommended. Revised lines 103: "…, each repetition entailing seven to eight feet of travel…"

P10, L133: your study did not observe nor report angles. Thus, the distinction should be clearly made to not mislead your readers. Please add a sentence to clarify this information.

Response: Revised as recommended. Revised lines 190-192: "Although this meta-analysis assessed joint angle reliability, while the current study inspected X-Y coordinate data, this lends further…"

P10, L135: define serial measures please for the lay reader.

Response: Revised as recommended. "test-retest" was added in parentheses on line 193.

P11, L141: is the Nyman study a white paper? If so, that needs to be specified and documented as such. If Nyman reports raw data, please report the numbers for direct comparisons.

Response: Yes, the Nyman study is a white paper. This was amended in the reference page as noted below. It does not report any raw data.
Nyman, E. (2017). A comprehensive evaluation of the trazer system : verification. [White paper]. 23 Consulting LLC.

P12, L159: or the vicon data was oversampled and thus more accurate at detecting max/peaks. Please adjust this wording.

Response: Revised as recommended. Revised lines 225-226: "…allows more accuracy in detecting peaks and maximums."

P12, L161-172: please update limitations to include sampling frequency mismatch.

Response: Revised as recommended, as noted above regarding lines 249-253.

P12, L177: for this final section, please provide the clinical relevance of having more reliability measures such as the TRAZER for its specific applications in sports and the like. This will help our non-specialized individuals.

Response: Revised as recommended. The following paragraph was added in lines 238-248. "*Clinical and Translational Impact.* A key advantage of the TRAZER system is the ability to closely replicate sport demands by presenting visual-cognitive virtual reality challenges that elicit quantifiable whole-body movement responses. Confidence in the system's measurements of reactive responses for clinical documentation of pre- and post-injury performance capabilities and assessment of injury risk is supported by some of the validity and reliability coefficients derived from our testing. Exceptionally close agreement of the total distance measurement derived from TRAZER with that from the Vicon system, along with very good test-retest reliability, clearly support its use as an indicator of whole-body movement precision in deactivation of virtual reality targets. These data indicate that TRAZER is appropriate for use as a baseline measure, coupled with post-injury quantification of return to pre-injury levels of functional capacity; thus allowing for more individualized return-to-play protocols."

Figure 2 to 4: they are rather blurry – can you provide more high resolution images or recreate using something like PRISM to have jpg images?

Response: Revised as recommended. New, cleaner, graphs were created in 300 dpi.

Reviewer #2,

Thank you for your thorough and thoughtful comments on our manuscript. We appreciate the opportunity to address the identified shortcomings, and feel that our manuscript is much stronger as a result. Our responses to each of your comments are detailed below. Additionally, changes within the manuscript have been highlighted for ease of location.

- Lines 7-9: First sentence is a bit of a run-on sentence; consider breaking into two sentences and linking them with a connecting phrase. (Ex: injury… As such, clinically…)

Response: Revised as recommended.

- Line 12: Consider addition of space requirement and/or software licenses to cons against motion capture

Response: Revised as recommended. Added "…and portable means of acquiring…" in abstract.

- Authors mention collecting maximum speed, but don't mention this in the methods (mention maximum velocity). As mentioned below under specific concerns for the Methods, please clarify which of these variables were collected as they mean different things.

Response: Velocity is the correct variable. This was adjusted in the abstract.

- No key words listed

Response: The following key words were added: reliability, validity, Kinect, non-immersive virtual reality, functional performance

- Lines 2-3: Consider expanding on what the functional movement assessment typically consists of (e.g., gait analysis)

Response: Thank you for this comment. The definition of functional movement assessments has been clarified. The second paragraph expands upon functional

movement assessments. With reviewer recommendations, the paragraph is now more detailed. Revised lines 13-31: "Lab-based three-dimensional (3D) motion capture systems are the gold standard in functional movement analysis, with reported excellent reliability ($ICC_{3,k} > .93$) (Ford, Myer, & Hewett, 2007) and validity (± 0.198 mm) (Vicon.com). However, they have limited clinical application due to financial, spatial, and temporal costs (Maykut, Taylor-Haas, Paterno, DiCesare, & Ford, 2015) and expertise needed to collect and interpret data. Other means of assessing quality of human movement have been developed and studied extensively, including the Functional Movement Screen (Cook, Burton, & Hoogenboom, 2014), Star Excursion Balance Test (Gribble, Hertel, Facsm, & Plisky, 2012), and the Landing Error Scoring System (Padua et al., 2009). These tools, however, lack ecological validity, as their required movements are anticipated, whereas responding to unanticipated events and simultaneous performance of cognitive and motor tasks are typically required during athletic activities. Unanticipated events can lead to sensory prediction errors and improper muscle co-contractions, potentially resulting in musculoskeletal injury. Identifying deficits in the simultaneous processing of environmental stimuli and task constraints and in the ability to preplan correct motor sequences (feed-forward) is important for injury prevention. Virtual reality systems that track body movements in response to visual stimuli may be valuable for assessing neuromechanical responsiveness, or the ability to optimally integrate neurocognitive and neuromuscular processes (Wilkerson et al., 2018), and integrated perception-motor neural processes (Teel, Gay, Johnson, & Slobounov, 2016), which appear to be crucial for preventing athletic injuries (Wilkerson et al., 2018)."

- Lines 2-6: Citation needed

Response: Citation added as recommended. (Dinc et al. 2017); revised line 5.

- Lines 7-9: This statement could have increase weight if the authors provide statistics on days lost to injury, health care costs, long term injury risk (e.g., knee OA)

Response: Revised as recommended. Lines 7-12: "Primary injuries, such as anterior cruciate ligament injuries, have been shown to precipitate long-term conditions such as osteoarthritis (Luc, Gribble, & Pietrosimone, 2014). Prevention of chronic conditions such as osteoarthritis could have a tremendous economic burden on the United States, which accounted for $11 billion in medical care costs and earnings losses (Brown, Johnston, Saltzman, Marsh, & Buckwalter, 2006)."

- Line 10: The first sentence can be omitted

Response: Stricken as recommended.

- Line 13: Citation needed

Response: Citation added as recommended. (Maykut et al. 2015); revised line 16-17.

- Lines 16-20: Feed forward mechanisms are important as well

Response: Thank you for this thoughtful comment. We agree that feed-forward mechanisms are important. We have added the following language in lines 25-29: "Identifying deficits in processing environmental stimuli and task constraints, along with the inability to preplan correct motor sequences (feed-forward) is important for injury prevention. Virtual reality systems that track body movements in response to visual stimuli may be valuable for assessing neuromechanical responsiveness, or the

ability to optimally integrate neurocognitive and neuromuscular processes (Wilkerson et al., 2018), and...”

- Line 18: I like this sentence as it makes a great point, however, it can be expanded upon. For instance, unanticipated events can lead to improper muscle tone and joint stiffness, leading to a sensory prediction error, and thus a musculoskeletal injury.

Response: Edited as recommended. Revised lines 22-31: “…whereas responding to unanticipated events and simultaneous performance of cognitive and motor tasks are typically required during athletic activities. Unanticipated events can lead to sensory prediction errors and improper muscle co-contractions, potentially resulting in musculoskeletal injury. Identifying deficits in the simultaneous processing of environmental stimuli and task constraints and in the ability to preplan correct motor sequences (feed-forward) is important for injury prevention. Virtual reality systems that track body movements in response to visual stimuli may be valuable for assessing neuromechanical responsiveness, or the ability to optimally integrate neurocognitive and neuromuscular processes (Wilkerson et al., 2018), and integrated perception-motor neural processes (Teel, Gay, Johnson, & Slobounov, 2016), which appear to be crucial for preventing athletic injuries (Wilkerson et al., 2018).

- Line 31: citation needed

Response: Citation added as recommended on line 31: (Wilkerson et al. 2018)

- Line 40: Avoid colloquial phrases like "on the other hand"; consider "conversely"

Response: Revised as recommended.

- The introduction should include more information on TRAZER (i.e., what does it measure, how, etc.)

Response: Edited as recommended. Language was added in the introduction to briefly describe the TRAZER. Further, an *Instrumentation* subsection was added in the Methods for more fully describe the TRAZER. Revised lines 44-50: “Briefly, the TRAZER system utilizes a Kinect camera. The system employs an infrared camera to create a two-dimensional (2D) representation on a video monitor. An individual interacts with the system by entering the capture field, approximately 1.75 x 1.75 meters. Once within the field, a simulated person appears on the monitor and responds to movement as the participant responds to visual targets randomly appearing on the perimeter of the capture area. All data captured by the TRAZER system is processed by imbedded proprietary algorithms, making it unique to a stand-alone Kinect camera.”

- Line 51: Please expand on "time and space constraints" which limited data collection

Response: This clause was omitted.

- Line 76: Operationally define the base of the spine

Response: Per TRAZER, this is what the system tracks. Upon email conversations, it was determined that S2 would be the best anatomical representation. Wording was adjusted in lines 119-120 to indicate this: “…”base of the spine”, which was operationalized as the digitized S2…”

- Line 78: Citation

Response: Citation added for the 'signal' R package in line 123. We also revised the citations in lines 123-124 to better match with the respective R packages.

- The authors do not justify their test-retest interval of 3 consecutive days of testing, as typically, there is a longer time interval between each assessment time. Some expansion on this might benefit the reader as a concern arises for a potential learning effect (although the results disagree with that; p=0.32).

Response: Revised as recommended. Revised lines 111-112: "Relying on pilot testing and anecdotal accounts, three consecutive days were used to account for a possible learning effect on the first day."

- Maximum speed is not mentioned a variable collected in the methods section (but maximum velocity is). Please clarify which of these variables were collected via TRAZER and 3D motion capture.

Response: Velocity is the correct variable. This has been corrected throughout.

- Statement needed on IRB approval in addition to existing informed consent statement

Response: Revised as recommended. Revised lines 87-89: "The reliability and validity components of this study were approved by the university's Institutional Review Board."

- A subsection called "Instrumentation" would be beneficial to the reader whereby the authors can further explain the TRAZER system. For example, how does it capture?

Response: Thank you for this helpful comment. In accordance with your (and other reviewers' recommendations, an *Instrumentation* subsection has been added in lines 90-106: "*Instrumentation.* TRAZER uses a depth-sensing Microsoft Kinect camera to create a three dimensional map of a 1.75 x 1.75 meter capture area. Anatomical landmarks (e.g., joint centers) are determined with a randomized decision forest algorithm with a one millisecond latency (Menna, Remondino, Battisti, & Nocerino, 2011; Nyman, 2017; Shotton et al., 2013). Specifically, each participant stands in the center of the capture area facing the TRAZER television screen for a brief (~5 second) calibration, during which the Kinect camera recognizes and identifies the participant. Following calibration, a visual target randomly appears at one of eight possible locations on the perimeter of the capture area (forward, backward, left, right, forward left diagonal, forward right diagonal, backward left diagonal, or backward right diagonal) (Figure 1). Once the indicator appears, the participant moves as quickly as possible to the location. Once TRAZER detects the participant in the correct location, the indicator disappears, and the participant returns to the start position to prepare for the next repetition. The protocol consists of forty repetitions (five at each of the eight possible locations), each entailing seven to eight feet of travel, and taking approximately three minutes to complete. TRAZER does not output raw data, but pre-defined performance metrics such as *reaction time, average/maximum velocity, average/maximum acceleration and deceleration, and total distance* traveled."

- Lines 118-123: These sentences can be more concise, all the while stating the same thing.

Response: Revised as recommended. Revised lines 177-179: "Our hypothesis pertaining to reliability was supported, in that reliability measures were rated as good to excellent. Our hypothesis pertaining to validity was not supported, in that concurrent validity was generally rated as poor."

- Lines 123-125: Please explain the translational and clinical relevance of this finding (as to align with the journal's mission)

Response: Revised as recommended. The following paragraph was added in lines 238-248. *Clinical and Translational Impact.* A key advantage of the TRAZER system is the ability to closely replicate sport demands by presenting visual-cognitive virtual reality challenges that elicit quantifiable whole-body movement responses. Confidence in the system's measurements of reactive responses for clinical documentation of pre- and post-injury performance capabilities and assessment of injury risk is supported by some of the validity and reliability coefficients derived from our testing. Exceptionally close agreement of the total distance measurement derived from TRAZER with that from the Vicon system, along with very good test-retest reliability, clearly support its use as an indicator of whole-body movement precision in deactivation of virtual reality targets. These data indicate that TRAZER is appropriate for use as a baseline measure, coupled with post-injury quantification of return to pre-injury levels of functional capacity; thus allowing for more individualized return-to-play protocols.

- Line 131-132: Should be Lopes's since it's a singular possessive noun

Response: Revised as recommended in lines 488-189.

- Lines 134-136: This is truly the take home message of this finding, as such, the reader would benefit from an expansion on these ideas and sentences.

Response: Thank you for the thoughtful comment. We have expanded this idea in lines 194-198 as follows: "For instance, periodic measurements could be obtained following a lower extremity injury to quantify the extent to which a patient achieves superior functional performance. Furthermore, the results of our study indicate that baseline TRAZER testing could be appropriate as a post-injury comparison to determine the point at which a patient reaches pre-injury level of function."

- Line 138: Define more affordable (how much more?)

Response: "Affordable" was replaced with "clinically feasible" throughout the manuscript. TRAZER is priced as a subscription, so cost is variable.

- Line 141: "Nyman, n.d.": No date? This is a M.S. Thesis project and the comparison isn't appropriate for a full paragraph relative to a conference abstract.

Response: We apologize for this key oversight. The date is 2017. It is a white paper. The year was placed in text and the reference updated, as noted in the comment below.

- Line 155-156: As mentioned in the methods, the fact that Vicon was sampled at 60 Hz seems like a poor reasoning in error, as the sampling rate could have easily been

Response: The discussion was supplemented to expand on this in lines 230-237 as noted in the comment below.

- The concept of TRAZER's proprietary algorithms should be expanded upon as I view this as a large deterrent from using TRAZER. Can the user trust this "black box" method? Additionally, the differences in methods/results between the study herein and Nyman's should be expanded upon further.

Response: Revised as recommended. Added the following to lines 212-219: "It should be noted that Nyman's work was a white paper, for which raw coordinate data were available. As such, TRAZER time series trajectories of the S2 marker were overlaid

on Vicon-obtained trajectories and then submitted to a validity analysis, whereas we assessed the validity of TRAZER-reported metrics with metrics calculated with Vicon coordinate data. Although Nyman reported excellent validity ICCs for time series data, coupling this with the current data suggests that gross metrics (e.g., trajectories, averages) are more valid than discrete (e.g., peaks, maximums, minimums) metrics."

- The difference in sample rates could have been avoided, why did the authors not change the Vicon system sampling rate to 30 Hz for ease of comparison? This seems like a study design oversight that warrants further discussion.

Response: We do agree that the difference in sampling rates likely accounts for some of the accuracy difference between systems. However, as a gold standard, 3D motion capture systems typically sample at higher frequencies, thus this was the comparison we chose to make. Given that the length of each trial was ~3 minutes, 60Hz was the highest frequency we felt comfortable would not overwhelm the system. Nevertheless, we do concede that this is a limitation and warrants further discussion. We have expanded the discussion in lines 230-237: "The authors elected for a 60 Hz Vicon sampling rate because this is more representative of a gold-standard motion capture collection. This is consistent with other researchers who have used different sampling frequencies when comparing two-dimensional with three-dimensional motion capture (Clark et al., 2012; Maykut et al., 2015; Nyman, 2017). In fact, in the white paper reported by Nyman, three dimensional motion capture was sampled at 120 Hz, while TRAZER was sampled at 30 Hz. Thus, although different sampling rates likely partially explain the systematic differences observed between the TRAZER and Vicon in the current study, it does not fully account for the discrepancy." We have also added this as a limitation in revised lines 249-253: "An important limitation of this study was the different sampling rates of each system; TRAZER captured at 30 Hz, while Vicon captured at 60 Hz. We acknowledge that this difference may partially account for discrepancies in accuracy. However, as gold standard 3-dimensional motion capture typically samples at higher frequencies, this is a more externally valid comparison."

- Were participants tested at the same time each day? If not, this could be added to the limitations as reaction time may differ at 8AM vs. 3PM.

Response: Yes, participants were tested at the same time each day. This was added in line 110 ("…at similar times each day") and reiterated in lines 254 ("…and were tested at similar times each day.")

- Again, in reference to the journal's mission, the discussion needs a clear clinical and/or translational aspect to the study

Response: Revised as recommended, per the comment above.

References:

- Line 226: Can the Nyman source be updated further? Is this a website, book, journal, op-ed?

Response: Yes, thank you for catching this oversight. The Nyman source is a white paper. The reference has been updated to read: Nyman, E. (2017). A COMPREHENSIVE EVALUATION OF THE TRAZER SYSTEM : VERIFICATION. [White paper]. 23 Consulting LLC."

- We there any power analysis conducted? (https://pubmed.ncbi.nlm.nih.gov/24197712/)

Response: We did not conduct an a priori power analysis. Lines 81 and 83 were revised to state that we recruited convenience samples.

- Figure 2: Crop graph to remove the bar on the left outer edge of the graph

Response: New, cleaner, graphs were created in higher resolution, without the bar on the left.

- As mentioned above under Abstract and Methods, there seems to be a confusion on variable name (maximum speed vs. maximum velocity); please clarify which of these were used in Table 2 and Figure 3.

Response: Thank you for your attention to this detail. Velocity is the correct variable. This was revised in Table 2.

Reviewer #3,

Thank you for your thorough and thoughtful comments on our manuscript. We appreciate the opportunity to address the identified shortcomings, and feel that our manuscript is much stronger as a result. Our responses to each of your comments are detailed below. Additionally, changes within the manuscript have been highlighted for ease of location.

Background:

1. Some additional background on the TRAZER system and its mechanism of kinematic data capture is needed. Kinect is mentioned here. How does TRAZER compare to the Kinect? The gap in research can't just be that this is a new system so we should test it also. There are a lot of new systems, why not one of them? TRAZER is chosen why? It is novel how?

Response: Revised as recommended. Language was added to explicitly state that the TRAZER system uses the Kinect camera, as well as further justification on TRAZER's novelty and why TRAZER was selected to examine, as noted in a comment below. Revised lines 44-50: "Briefly, the TRAZER system utilizes a Kinect camera. The system employs an infrared camera to create a two-dimensional (2D) representation on a video monitor. An individual interacts with the system by entering the capture field, approximately 1.75 x 1.75 meters. Once within the field, a simulated person (avatar) appears on the monitor and mirrors the participant as they respond to visual targets randomly appearing on the perimeter of the capture area. All data captured by the TRAZER system is processed by embedded proprietary algorithms, making it unique to a stand-alone Kinect camera."

2. Prior work on the Kinect was appropriately mentioned in the background. It may be helpful to state why Kinect has "variable test-retest reliability." Was the Kinect's variable reliability related to which kinematic variables were measured? If reliability of measurement may be a limitation of the kinect in some settings or for some applications, as is implied here, what are the potential advantages of TRAZER that may make it theoretically more reliable than the Kinect (if any)?

Response: Revised as recommended to include TRAZER vs. Kinect differences and clarifying prior reliability data. Revised lines 50-59: "The Kinect camera has been evaluated for reliability and validity compared to 3D motion capture systems during postural control and balance tasks (Clark et al., 2012, 2015; Eltoukhy, Kuenze, Oh, & Signorile, 2018), dynamic side-cut maneuvers (Eltoukhy et al., 2019), squatting (Schmitz et al., 2015), and single leg squat (Ressman, Rasmussen-Barr, Grooten, 2020). Studies report good to excellent concurrent validity for kinematic angles during a postural balance task (Pearson's $r > 0.90$) (Clark et al., 2012), (ICC > 0.75)

(Eltoukhy, Kuenze, Oh, & Signorile, 2018), side-cutting maneuvers (absolute agreement ICC range = .767-.989) (Eltoukhy et al., 2019), and squatting (Pearson's r > 0.55) (Schmitz et al., 2015). Test-retest reliability has variable results with excellent reliability (ICC > 0.90) reported by Schmitz et al., 2015, to modest reliability (ICC ≥ 0.70) reported by Clark et al., 2015."

Method:

1. "Due to time and space constraints". Be more specific, what time and space constraints? Otherwise, omit.

Response: This phrase was omitted.

2. "participants were not digitized" What does that mean?

Response: Revised to read in lines 109: "…with any 3D markers or instrumentation."

3. It is unclear from the write-up what data points the TRAZER system provided. Raw coordinates of the S2 joint only?

Response: The TRAZER system outputs the pre-defined metrics of total distance, maximum and average acceleration, reaction time, etc. Unfortunately, it does not output coordinate data. A brief subsection within the methods was added ("*Instrumentation"*) to more explicitly explain the TRAZER system. Revised lines 90-106: "*Instrumentation.* TRAZER uses a depth-sensing Microsoft Kinect camera to create a three dimensional map of a 1.75 x 1.75 meter capture area. Anatomical landmarks (e.g., joint centers) are determined with a randomized decision forest algorithm with a one millisecond latency (Menna, Remondino, Battisti, & Nocerino, 2011; Nyman, 2017; Shotton et al., 2013). Specifically, each participant stands in the center of the capture area facing the TRAZER television screen for a brief (~5 second) calibration, during which the Kinect camera recognizes and identifies the participant. Following calibration, a visual target randomly appears at one of eight possible locations on the perimeter of the capture area (forward, backward, left, right, forward left diagonal, forward right diagonal, backward left diagonal, or backward right diagonal) (Figure 1). Once the indicator appears, the participant moves as quickly as possible to the location. Once TRAZER detects the participant in the correct location, the indicator disappears, and the participant returns to the start position to prepare for the next repetition. The protocol consists of forty repetitions (five at each of the eight possible locations), each entailing seven to eight feet of travel, and taking approximately three minutes to complete. TRAZER does not output raw data, but pre-defined performance metrics such as *reaction time, average/maximum velocity, average/maximum acceleration and deceleration, and total distance* traveled."

4. If TRAZER calculations are proprietary, how can it be assured that the same calculations are being used by the TRAZER system and for the Vicon data?

Response: We appreciate your concern on this matter. It is for this reason that we decided not to calculate average velocity and acceleration for validation purposes, as we were unsure how TRAZER handled velocity and acceleration vectors. The three variables we did elect to include in the validity analysis (total distance, max velocity, max acceleration) are straightforward and we felt confident in identifying maximum points in a time series.

5. Did the researchers calculate the metrics for both datasets from available outputs of the TRAZER system?

Response: We did not, as TRAZER only outputs pre-defined metrics and we could not access raw data. The addition of the *Instrumentation* subsection noted above also addresses this comment.

6. The clinical relevance of each of the outcomes of interest should be explained, e.g., why would total distance be of interest to a clinician? E.g., joint angles such as knee flexion during a squat have more obvious relevance than the speed and location of a particular body joint. The clinical rationale needs to be clearer in both the Intro and the Methods section.

Response: Revised as recommended. Thank you for this thoughtful comment. In brief, we believe these measures are useful because they are indicative of overall functional performance and one's ability to return to sport, whereas more process-oriented measures, such as knee flexion angle, would be more appropriate for early stage rehabilitation. Revised lines 39-43: "Its novelty lay in its output metrics. Namely, the measures of *total distance*, *maximum velocity*, *maximum acceleration*, and *reaction time* are indicative of overall functional performance and as such, may be useful for clinicians and coaches aiming to determine one's ability to return to sport or their improvement in overall function."

Results:

1. The Bland Altman plots seem to show a proportional bias, i.e. that the errors are proportional so that the larger values have more error compared to Vicon. It is presumed that the measurements of the TRAZER system are on the x-axis, but this should be stated.

Response: The x-axis in Figures 2-4 is the mean of TRAZER and Vicon. The axis labels were revised in each of the three graphs to reflect this.

Discussion: The utility of the TRAZER system should be contrasted with other available systems for which reliability and validity have been measured, e.g., Kinect.

Response: The Introduction was clarified to explicitly state that the TRAZER uses the Kinect camera. Reliability and validity numbers for the Kinect have been added in the introduction.

Reviewer #4,

Thank you for your thorough and thoughtful comments on our manuscript. We appreciate the opportunity to address the identified shortcomings, and feel that our manuscript is much stronger as a result. Our responses to each of your comments are detailed below. Additionally, changes within the manuscript have been highlighted for ease of location.

The introduction must be assertive towards the Trazer System. The first two paragraph does not add much to the rationale of the study. The authors could bring technical details of the system and studies done with it, for instance, to demonstrate the necessity of the present study.

Response: The introduction was heavily revised to justify the need for the study and to include further details regarding the TRAZER system and existing evidence on the Kinect camera it uses. Additionally, we refer you to the addition of an *Instrumentation* subsection, which we are hopeful will address several of your reservations.

The third paragraph is too long. It can be broken in two and the authors should concentrate on it to improve the rationale of the study. For instance, the explanation of how the Trazer System tracks and reconstruct movements within the cartesian plane would be relevant. A number of processes of the way the cameras of the Trazer system capture body's motion and

the algorithm reconstructs it in bi-dimensional or tri-dimensional perspectives are relevant to understand the comparisons with Vicon system.

Response: Revised as suggested. Third paragraph broken up and additional information on the TRAZER system added. *Instrumentation* subsection (lines 90-106) added that addresses this comment as well. Revised lines 44-50: "Briefly, the TRAZER system utilizes a Kinect camera. The system employs an infrared camera to create a two-dimensional (2D) representation on a video monitor. An individual interacts with the system by entering the capture field, approximately 1.75 x 1.75 meters. Once within the field, a simulated person (avatar) appears on the monitor and mirrors the participant as they respond to visual targets randomly appearing on the perimeter of the capture area. All data captured by the TRAZER system is processed by embedded proprietary algorithms, making it unique to a stand-alone Kinect camera."

With respect to the rationale of the study, it is reasonable to bring details of each system, so we will be aware of the differences of each system.

Response: Revised as suggested. Revised lines 74-78: "If TRAZER demonstrates agreement against the gold standard 3D motion capture, the utilization of a more clinically-feasible alternative may expand clinicians' resources. The ability to accurately measure movement via a commercially available system has the potential for widespread adoption. Reliability is important in the clinic as well as in the laboratory, as reproducing consistent outcomes allow for accurate comparisons of patient progress."

To better suit a wide range of interested reader I suggest to explain the theoretical meaning of reliability and validity. With respect to the reliability testing is not clear why that is relevant? The manner in which reliability testing was designed in the present study is not a straightforward measure of it. Actually, the protocol tested differences in the measures of performing the task with the Trazer system in three different days. The validity testing is the relevant to test the Trazer system with an accurate system.

Response: Language was added throughout the Introduction to describe how assessing reliability of the TRAZER system will benefit clinicians and coaches seeking to determine one's ability to return to functional capacity or to assess improvement in sport-related function. Furthermore, reliability and validity were operationally defined in lines 108 and 113, respectively.

The procedures of data collection need more details and better organized. Separate the explanations of the two experiments in different paragraphs or topics. Provide details of the procedures for both experiments (reliability and validity).

Response: Revised as recommended. Validity was created as a new paragraph and greater details provided throughout.

Operational definition of reliability and validity should be included in procedures.

Response: Revised as recommended. Revised lines 108: "…reliability, operationally defined as test-retest consistency…" and lines 113: "…validity, operationally defined as absolute agreement…"

The sentence on line 59 "to obtain test-retest reliability data, participants were not digitized." does not make sense without a proper explanation.

Response: Revised to read in lines 109: "…with any 3D markers or instrumentation."

How many trials or repetitions did the participants perform in each day for the first experiment (i.e., reliability)? Was the order of repetition random? Was there any procedure to calibrate each participant's body?

Response: Forty repetitions were performed on each day. Each repetition required 7-8 feet of travel and the 40-rep protocol took approximately 3 minutes. Yes, the order of repetition was randomly determined by the TRAZER system. There was a brief calibration process, as detailed in revised lines 94-96: "Prior to the protocol, each participant stands in the center of the capture area facing the TRAZER television screen for a brief (~5 second) calibration, during which the Kinect camera recognizes and identifies the participant." The following language was also added to lines 102-104: "The protocol consists of forty repetitions (five at each of the eight possible locations), each repetition entailing seven to eight feet of travel, with a complete trial taking approximately three minutes to complete."

Explain the technical aspects of data collection with the Trazer system. How does the system recognize a human body and its movement per segment? Does the Trazer system reconstruct the movement in bi-dimensional or tri-dimensional perspective?

Response: Revised as recommended. An *Instrumentation* subsection was added in lines 90-106: *Instrumentation.* TRAZER uses a depth-sensing Microsoft Kinect camera to create a three dimensional map of a 1.75 x 1.75 meter capture area. Anatomical landmarks (e.g., joint centers) are determined with a randomized decision forest algorithm with a one millisecond latency (Menna, Remondino, Battisti, & Nocerino, 2011; Nyman, 2017; Shotton et al., 2013). Specifically, each participant stands in the center of the capture area facing the TRAZER television screen for a brief (~5 second) calibration, during which the Kinect camera recognizes and identifies the participant. Following calibration, a visual target randomly appears at one of eight possible locations on the perimeter of the capture area (forward, backward, left, right, forward left diagonal, forward right diagonal, backward left diagonal, or backward right diagonal) (Figure 1). Once the indicator appears, the participant moves as quickly as possible to the location. Once TRAZER detects the participant in the correct location, the indicator disappears, and the participant returns to the start position to prepare for the next repetition. The protocol consists of forty repetitions (five at each of the eight possible locations), each entailing seven to eight feet of travel, and taking approximately three minutes to complete. TRAZER does not output raw data, but pre-defined performance metrics such as *reaction time, average/maximum velocity, average/maximum acceleration and deceleration, and total distance* traveled."

Start the description for the validity procedures in a new paragraph to separate from the reliability experiment.

Response: Revised as recommended.

Line 71, the sentence "…they were digitized" needs a concise and clear explanation.

Response: Revised as recommended. Revised lines 114-115: "…in which a reflective marker placed on the S2 spinous process was digitized with Vicon Nexus."

Explain whether the Vicon system and Trazer system are tracking the exact same point on the body? And how are they tracking the points according to their cameras?

Response: TRAZER reports that they track the "base of the spine." Vicon tracked the retroreflective marker placed on S2, our operational definition of the base of the spine. Building on Kinect technology, TRAZER uses decision forest algorithm to identify anatomical landmarks, as noted in the added *Instrumentation* subsection.

What is the accuracy of each camera's system (Trazer and Vicon)?

Response: Vicon's reported accuracy is ± 0.098 mm (Vicon.com). Discrete joint kinematics and kinetics for landing variables have been determined as reliable (ICC range = .93-.95) (Ford, Myer, & Hewett, 2007). Revised lines 14-15: "…with reported excellent reliability ($ICC_{3,k} > .93$) (Ford, Myer, & Hewett, 2007) and validity (± 0.198 mm) (Vicon.com)."

Although the data reduction from Trazer System is built in and apparently there is no possibility the authors treat the raw data, then the authors must describe the process of treatment of raw data by the built-in algorithm.

Response: Unfortunately, the treatment of raw data by the built-in algorithm is also proprietary information to which we are not privy. To supplement this shortcoming, the instrumentation subsection was added.

Discussion section needs to improve in a way that the authors must demonstrate differences that each system may produce as a result of data treatment.

Response: Revised as recommended. Specifically, revised lines 212-219 discuss the difference between assessing trajectories and discrete points, while revised lines 230-237 discuss further the implications of different sampling frequencies used by TRAZER and Vicon.

On line 155 the authors may suspect about the differences in sampling rate between both systems. That is probably the reason of differences between the systems. However, a detailed of data treatment and the algorithm used by each system to calculate each variable would provide a clue of the lack of correlation. For instance, how each system calculated each variable? If Trazer has one camera and Vicon has 8 cameras.

Response: As noted above, TRAZER (Kinect) uses a randomized decision forest algorithm to identify anatomical landmarks using a single camera. With eight cameras, Vicon uses a least squares algorithm to identify points in three-dimensional space. Unfortunately, further information on data treatment by TRAZER is not available.

The discussion is confusing. For example, '…high absolute agreement…' (line 137), that is not adequate to say when the ICC was carried out and not Cohen's Kappa.

Response: ICCs can reflect absolute agreement, as supported by the references below.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, *1*(1), 30.

Line 155: what the authors want to say with lack of congruity? Were the results about congruity presented in the results section?

Response: By "congruity," we meant agreement. "Agreement" was substituted for "congruity" in lines 199 and 221.

I did not agree that motivation of the participants would be a limitation of the study. Actually, the limitation is on the procedure or technique used by the authors to examine the reliability of the system.

Response: Revised as recommended. The limitation of participant motivation was removed and the limitation of different sampling frequencies was added. Revised lines 249-253: "An important limitation of this study was the different sampling rates of each system; TRAZER captured at 30 Hz, while Vicon captured at 60 Hz. We acknowledge that this difference may partially account for discrepancies in accuracy. However, as gold standard 3-dimensional motion capture typically samples at higher frequencies, this is a more externally valid comparison."

2nd Editorial decision
06-Dec-2020

Ref.: Ms. No. JCTRes-D-20-00088R1
Reliability and Concurrent Validity of TRAZER Compared to 3-Dimensional Motion Capture
Journal of Clinical and Translational Research

Dear Dr. Hogg,

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 Jan 05, 2021.

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: Dear author

Well done on the comments/clarifications. All of my comments/concerns have been met at this time.

Reviewer #2: Thank you for asking me to review this revision. While the authors have taken strong measures to improve their manuscript through the various reviewers' feedbacks, the manuscript is still not ready for publication in JCTRes as it lacks flow and succinct storytelling. It would behoove the authors to further revise their manuscript so it reads more concisely, without straying to far off topic, and double check their manuscript to avoid repetitiveness of concepts.

Abstract:
- Key words should not include words that are in the title; they should be different, yet descriptive so when someone is searching PubMed or even Google your article will appear. Consider changing or omitting key words "reliability" and "validity" as these are in your title.
Introduction:
- Lines 6-12: While these are interesting data and facts and the authors have taken note of my previous suggestion it was simply an example. It appears as the information presented is now going beyond the scope of this article. Yes, injury prevention is important as injuries can lead to further dysfunction, but it appears the authors have "gone down a rabbit hole". The authors should consider revising this again to be more concise (e.g., functional movement screens are used to find dysfunction ◊ dysfunction is linked to further injury which can lead to morbidity and dysfunction in later life, causing a social and economic burden). Keep it broad if possible so as not to get too specific.
o First paragraph should simply highlight what the "problem" is. So the problem here (I assume) is that injuries happen. They authors should decide what component they are to focus on here (injuries happen due to premature return to sport? Injuries happen because failure to identify movement dysfunction?) and adjust this first paragraph accordingly, because as it stands it lacks direction.
- Line 25 & 27: Citation needed. Consider:
o Swanik CB, Covassin T, Stearne DJ, Schatz P. The relationship between neurocognitive function and noncontact anterior cruciate ligament injuries. Am J Sports Med. 2007;35(6):943-8.
o Howell DR, Lynall RC, Buckley TA, Herman DC. Neuromuscular Control Deficits and the Risk of Subsequent Injury after a Concussion: A Scoping Review. Sport Med [Internet]. 2018 Aug 13;48(5):1097-115. Available from: https://doi.org/10.1007/s40279-018-0871-y
o McPherson AL, Nagai T, Webster KE, Hewett TE. Musculoskeletal injury risk after sport-related concussion: a systematic review and meta-analysis. Am J Sports Med. 2018;3:363546518785901.
- Line 43: citation needed
- Lines 40-59: This is all valuable information regarding the TRAZER, but some of it is more appropriate for the instrumentation section. Be brief yet descriptive here. The description of TRAZER as a non-immersive virtual reality system is needed along with a brief description (utilizes a 2d representation on a video monitor whereby a participant responds to stimuli and is recorded using a Kinect. It is reliable and valid for postural control and balance tasks.)
- The background on the Nyman white paper is much needed and welcomed addition to the introduction and adds to your manuscript's justification. The authors can still further revise this section to be more concise and link the prior paragraph better.
Methods:

- Lines 82-83: Thank you for include male/female breakdowns, however "Men/Women" seems to be more appropriate for this age group (Men/Women are nouns, male/female are genders). Additionally, the authors only need to list one. If there at 18 subjects (11 Women), the reader can infer there are 7 men.
- The addition of the Instrumentation section is very helpful and betters this manuscript.
- Lines 90-91: Repetitive of the introduction. Consider omitting this part from the introduction and keeping this information here in instrumentation.
- Line 106: Citation needed for background information on TRAZER
- Line 111: The plural of "anecdote" is not data, therefore the authors should not use anecdotal data to justify the test-retest interval. However, if the pilot data justified the interval timeframe (i.e., no learning effect, etc.) then that should suffice as a justification.
- Lines 117-118: Why were the data sampled at different frequencies? Vicon can be adjusted to match the sampling frequency of TRAZER. I see this was expanded upon in the discussion, but a brief statement might benefit the reader here in the Methods (e.g., consistent with previous reliability studies….)
- Line 138: Authors already operationally defined reliability so this sentence can be omitted

Results:
- Lines 159-161: This sentence is confusing to read. Do the authors mean with the exception of reaction time….. or with the exception of "reaction time, average deceleration, and maximum acceleration". Which variables are included in the exception?

Discussion:
- The first two sentences read poorly and staggered. The authors can take my suggestion or ignore it based upon their preferences, however it may read better if the authors wrote:
  o "We hypothesized that ……." "Our hypothesis was partially supported evidenced by good to excellent reliability measures, but poor concurrent validity" or along those lines.

Tables & Figures
- The addition of the two tables is extremely helpful.

Authors' response

December 12, 2020

Dear reviewer,

Thank you for your continued interest in improving our manuscript. After addressing your remaining concerns, we feel that our manuscript is more succinct and concise. Our line-by-line responses to your recommendations are detailed below. The revised manuscript is also highlighted in the appropriate locations for easy identification of changes. Thank you again for your thoughtfulness and thoroughness.

Abstract:
- Key words should not include words that are in the title; they should be different, yet descriptive so when someone is searching PubMed or even Google your article will appear. Consider changing or omitting key words "reliability" and "validity" as these are in your title.

  Response: These two key words were omitted and the term "reactive agility" was added.

Introduction:
- Lines 6-12: While these are interesting data and facts and the authors have taken note of my previous suggestion it was simply an example. It appears as the information presented is now going beyond the scope of this article. Yes, injury prevention is important as injuries can lead

to further dysfunction, but it appears the authors have "gone down a rabbit hole". The authors should consider revising this again to be more concise (e.g., functional movement screens are used to find dysfunction ◊ dysfunction is linked to further injury which can lead to morbidity and dysfunction in later life, causing a social and economic burden). Keep it broad if possible so as not to get too specific.
o First paragraph should simply highlight what the "problem" is. So the problem here (I assume) is that injuries happen. They authors should decide what component they are to focus on here (injuries happen due to premature return to sport? Injuries happen because failure to identify movement dysfunction?) and adjust this first paragraph accordingly, because as it stands it lacks direction.

Response: Revised as recommended. First paragraph was shortened to avoid mention of discrete injuries. Lines 6-12 were replaced with the following, more broad, verbiage in revised lines 7-9: "Portable and convenient acquisition of accurate movement assessment data is necessary for effective injury risk identification and ultimate prevention of sport-related injury."

- Line 25 & 27: Citation needed. Consider:
o Swanik CB, Covassin T, Stearne DJ, Schatz P. The relationship between neurocognitive function and noncontact anterior cruciate ligament injuries. Am J Sports Med. 2007;35(6):943-8.
o Howell DR, Lynall RC, Buckley TA, Herman DC. Neuromuscular Control Deficits and the Risk of Subsequent Injury after a Concussion: A Scoping Review. Sport Med [Internet]. 2018 Aug 13;48(5):1097-115. Available from: https://doi.org/10.1007/s40279-018-0871-y
o McPherson AL, Nagai T, Webster KE, Hewett TE. Musculoskeletal injury risk after sport-related concussion: a systematic review and meta-analysis. Am J Sports Med. 2018;3:363546518785901.

Response: Revised as recommended. All three citations were included in lines 24-25.

- Line 43: citation needed

Response: Revised as recommended. The following citations were added to revised lines 42-43: "(Wilkerson, Nabhan, & Crane, 2020; Wilkerson et al., 2018)."

- Lines 40-59: This is all valuable information regarding the TRAZER, but some of it is more appropriate for the instrumentation section. Be brief yet descriptive here. The description of TRAZER as a non-immersive virtual reality system is needed along with a brief description (utilizes a 2d representation on a video monitor whereby a participant responds to stimuli and is recorded using a Kinect. It is reliable and valid for postural control and balance tasks.)

Response: Revised as recommended. Lines 44-47 were shortened and some of this information was moved to the *Instrumentation* subsection, as noted below. Revised lines 46-49: "Briefly, the TRAZER system is a non-immersive virtual reality system that utilizes a Kinect camera. The system employs an infrared camera to create a two-dimensional (2D) representation on a video monitor whereby a participant responds to visual stimuli and is recorded by a Kinect camera."

- The background on the Nyman white paper is much needed and welcomed addition to the introduction and adds to your manuscript's justification. The authors can still further revise this section to be more concise and link the prior paragraph better.

Response: Revised as recommended. A transition sentence was added in the paragraph prior, and the paragraph on the Nyman white paper was revised for clarity and conciseness. Revised lines 58-60: "Although Kinect systems have been assessed for

accuracy, the accuracy of proprietary algorithms overlaid on Kinect, such as those provided by TRAZER, has yet to be determined."

Methods:
- Lines 82-83: Thank you for include male/female breakdowns, however "Men/Women" seems to be more appropriate for this age group (Men/Women are nouns, male/female are genders). Additionally, the authors only need to list one. If there at 18 subjects (11 Women), the reader can infer there are 7 men.

Response: Thank you for your attention to this detail. However, man/woman is gender and male/female is biological sex, and as such is the correct verbiage. We did revise to only list female, as recommended.

- The addition of the Instrumentation section is very helpful and betters this manuscript.

Response: Thank you for this comment. The authors agree that this was a much needed section and betters the manuscript.

- Lines 90-91: Repetitive of the introduction. Consider omitting this part from the introduction and keeping this information here in instrumentation.

Response: Revised as recommended. Omitted this redundant material from introduction. Also, revised lines 91-93: "Once within the field of view, a simulated person (avatar) appears on the monitor and mirrors the participant as they respond to visual targets randomly appearing on the perimeter of the capture area."

- Line 106: Citation needed for background information on TRAZER

Response: Revised as recommended. Revised lines 108-109: "(see Nyman, 2017a and https://trazer.com/science for comprehensive background information on TRAZER)."

- Line 111: The plural of "anecdote" is not data, therefore the authors should not use anecdotal data to justify the test-retest interval. However, if the pilot data justified the interval timeframe (i.e., no learning effect, etc.) then that should suffice as a justification.

Response: Revised as recommended. Mention of "anecdotal accounts" was deleted.

- Lines 117-118: Why were the data sampled at different frequencies? Vicon can be adjusted to match the sampling frequency of TRAZER. I see this was expanded upon in the discussion, but a brief statement might benefit the reader here in the Methods (e.g., consistent with previous reliability studies….)

Response: Revised as recommended. Revised lines 121-122: "…, consistent with previous validity studies that have used higher sampling rates for 3D motion capture (Clark et al., 2012; Maykut et al., 2015; Nyman, 2017a)."

- Line 138: Authors already operationally defined reliability so this sentence can be omitted

Response: Revised as recommended. This sentence was deleted.

Results:
- Lines 159-161: This sentence is confusing to read. Do the authors mean with the exception of reaction time….. or with the exception of "reaction time, average deceleration, and maximum acceleration". Which variables are included in the exception?

Response: All three variables are included in the exception. This was clarified in line 162: "With the exception of the three variables of…."

Discussion:
- The first two sentences read poorly and staggered. The authors can take my suggestion or ignore it based upon their preferences, however it may read better if the authors wrote:
o "We hypothesized that ……." "Our hypothesis was partially supported evidenced by good to excellent reliability measures, but poor concurrent validity" or along those lines.

Response: Thank you for this suggestion. We have revised as recommended. Revised lines 180-182: "We hypothesized that TRAZER would display moderate test-retest reliability and good to excellent concurrent validity. Our hypothesis was partially supported as evidenced by good to excellent reliability, but poor concurrent validity."

Tables & Figures
- The addition of the two tables is extremely helpful

Response: Thank you for this kind comment.

3rd Editorial decision
13-Dec-2020

Ref.: Ms. No. JCTRes-D-20-00088R2
Reliability and Concurrent Validity of TRAZER Compared to 3-Dimensional Motion Capture
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

Comments from the editors and reviewers: