Author’s response to reviews

Title: Reliability and validity of a Novel Kinect-based Software Program for Measuring a Single Leg Squat.

Authors:

John Ressman (john.ressman@ki.se)

Eva Rasmussen-Barr (eva.rasmussen.barr@ki.se)

Wilhelmus Grooten (wim.grooten@ki.se)

Version: 1 Date: 13 Feb 2020

Author’s response to reviews:

Stockholm 2020-02-13

Dear editor-in-chief,

We hereby resubmit our manuscript “Reliability and validity of a novel Kinect-based software program for measuring a Single Leg Squat”.

We would like to express our gratitude to the editor and to the reviewers for taking time to review our manuscript. We find all comments excellent and important. We have revised the manuscript accordingly and find that the manuscript has improved due to the comments. We sincerely hope that you will find the manuscript improved and more focused.

The material submitted is original and is not being submitted for publication elsewhere; either in whole or in part. There is no financial or other relationship that might be perceived as leading to a conflict of interest.

Kind regards,

Corresponding author:

John Ressman, PhD candidate
Department of Neurobiology, Care Sciences and Society | Division of Physiotherapy
Karolinska Institutet | Alfred Nobels Allé 23 | 141 83 Huddinge
+46 709 656076 | john.ressman@ki.se

Eva Rasmussen-Barr PhD and Associate Professor

Wim JA Grooten, PhD and Associate Professor
Dear editor and reviewers,

Thank you for the opportunity to revise our manuscript and to respond to the comments of the reviewers. We find that the comments are very reasonable and are willing to make alterations accordingly. In fact, these comments made a great improvement of the paper and we are very happy about it.

The point-by-point responses are given below.

1. Reviewer reports: Clint Hansen (Reviewer 1):
I only have one minor remark with regards to competing interests. At least one of the authors "Wim" appears in the youtube-video of the company (https://www.youtube.com/watch?v=CEE5nH6kn9M). The content is rather trivial but I would urge to declare somehow that you are not affiliated with the company or at least declare what kind of work or development you have done for them. Also I would urge you to declare if you have received the system by the company e.g. as a loan or if you have bought the system just to avoid any issues.
Kind regards from Kiel, Clint Hansen

We thank the reviewer for pointing this out. This is not a trivial issue and we will of course declare this in the manuscript as none of us are developers of the QinematicTM system nor affiliated to the company. The used QinematicTM system is owned by the Karolinska Institutet as Skandia Insurance supplied financial support to obtain the QinematicTM system to a previous paper [1]. In this previous paper, we were asked by Skandia Insurance to assess the validity and reliability of the system (posture, balance and side-bending capacity) and we found poor results for these tests.

Amendment at competing interest: page 30 line 2-3.
“The authors declare that they are not system developers of the QinematicTM system nor affiliated to the company, and there is no further competing interest whatsoever.”

2. A Kody Campbell (Reviewer 2):
Overall Impressions:
A major concern I have is if the ICC's were run with non-normal data. The authors' presented summary data using medians and IQRs and used non-parametric analyses for testing single leg squat measures from the Kinect based software between test sessions. The authors, however, do not indicate if the data were non-normally distributed and ran the ICC analyses on the data without indicating if the data had been transformed to satisfy normality assumptions. Without this critical information it is difficult to interpret the ICCs, SEMs and SDCs in this study. This needs to be addressed by the authors on any resubmission along with the correct reliability analyses for normal, non-normal, or transformed data.
We thank the reviewer for this important remark. We certainly understand the difficulties in interpreting the data without any assumptions of the normality distribution of the data. Indeed, we had the same discussion and thoughts in the first paper of QinematicTM where we explored the reliability and validity of posture, balance and side bending [1].
When the data were tested for normality of distribution, two out of eight variables were seen as non-normally distributed (left knee down and right knee up at occasion one) and to be certain not to overestimate our data, descriptive statistics were presented with medians and interquartile range (IQR) and non-parametric analyses were used to test for potential differences between the test occasions. Perhaps we were a bit too conservative in this regard and we could instead have used parametric tests for these six variables, together with non-parametric tests for the above-mentioned variables. However, in order to keep the paper readable, we decided not to use two different tests.

The main reason for calculating the ICC for these two non-normally distributed data and not transforming the data was that the SEM and MDC should be expressed in their original units for clinical use.

Moreover, we compared the ICC’s with non-parametric correlations (Spearman correlation coefficients), and the results showed that the results were not altered to a large extent. These results are now added in Table 2, together with a better description of this in the method section. The results indicate clearly that the non-normally distribution in these two variables did not influence the results in such a way that our conclusions lose their validity.

Amendment in the method section: page 13 line 3-7, 17-20.
“...To check for normality, means and medians were compared, together with visual analyses of histograms and distributional diagnostic plots. We also tested for skewness and kurtosis [2]. Two out of eight variables showed non-normally distributed data (left knee down and right knee up at occasion one) and in order not to overestimate our data, medians, interquartile range (IQR) and non-parametric statistics were used for the descriptive statistics.”
“...Since all variables were not-normally distributed, Spearman correlation coefficients were in addition calculated. The Spearman correlation coefficient was interpreted as; less than 0.3 low correlation, 0.3–0.5 fair correlation, 0.6–0.8 moderately strong correlation, at least 0.8 very strong [3, 4].”

Amendment in table 2: A column with Spearman correlation coefficient statistics are added.

Amendment in the result section: page 15 line 21-23, page 16 line 1-2 and line 4-6.
“...Since two variables were not-normally distributed (left knee down and right knee up at occasion one), Spearman correlation coefficients and in addition ICC were calculated for the relative reliability.”

“In comparison, Spearman correlation coefficient for the same three variables reached “moderately strong” correlation ranging from 0.61 to 0.68.”

“For Spearman correlation coefficient this variable reached a “fair correlation” of 0.53, which differ slightly compared to the ICC measurement in which a moderate correlation was found.”
2.B
The authors came to the correct conclusions based on the results of their study and don’t recommend using the Kinect based software for assessing frontal knee angles while performing a single leg squat. According to their construct validity analysis, when someone performs a single leg squat and their knee is over their foot the software has a 94% chance of correctly identifying that movement. In my opinion the software does not add anything substantial to a clinical assessment of a single leg squat done by a clinician. More of these points or a section of this nature should appear in the discussion (see specific comments for the authors). Otherwise, I was quite satisfied and impressed with the execution and presentation of this data.

See point 34 for an answer and an amendment.

3.
Abstract Line 21: Change cannot to should not

We thank the reviewer for this comment, it has now been corrected.

Amendment in abstract: page 2 line 21
“should not”

4.
Background: The background could be separated into 3 paragraphs. Split the first paragraph where it introduces objective measures of movement quality. This new second paragraph could be bolstered with further information on the need for refined and reliable measures of movement quality in order to assess injury risk.

We thank the reviewer for this proposal as the amendments have made the background better and more comprehensive. The first amendment is made at page 5 line 2-8 in the first paragraph and aim to make a smooth transition to the new second paragraph.

Amendment in background: page 5 line 2-8.
“An increased dynamic knee valgus and its interlinked malalignments of the lower extremity are proposed to be associated with overuse syndromes such as patellofermoral pain syndrome [1], iliotibial pain syndrome [2], femuro-acetabular impingement [3], tibial stress fractures [4] and injuries such as anterior cruciate ligament injuries [5]. To date, no cut-off point is established for when the degree of medial displacement of the knee, i.e. dynamic knee valgus, is to be considered a risk for these syndromes or injuries.”

Amendment in the background: page 5 line 13-25 and page 6 line 1-9.
“Several screening systems are proposed to observe injured athletes or for proactive purposes: the Functional Movement Screen (FMS) [6-9], Functional Movement Test +9 [10-12] and Test for Substitution Patterns (TSP) [13], but also used in single tests/single movements [14, 15]. A critical step in injury prevention is to determine the causes of injury which includes the understanding why an athlete may be at a greater risk in a given situation (risk factor) and how injury happens (injury mechanism/inciting event) [16]. As the causes of injury are multifactorial, comprehensive models to understand injury causations have been developed which emphasis intrinsic-, and extrinsic risk factors together with a careful description of the injury (injury mechanism/inciting event) [16, 17]. Intrinsic factors are, among others, described as modifiable physiological factors [16, 18, 19] such as dynamic
knee valgus which can be visually assessed in different functional test such as SLS, drop jump, vertical drop jump, lunge, one leg hop or crossover hop [20, 21]. Athletic screening for risk factors in order to predict or prevent injury has been criticized since they cannot predict injury with sufficient accuracy. Even though some authors have found a highly significant correlation between dynamic knee valgus and injury risk, it is impossible to predict future injuries based on the results of these tests [22]. Verhagen et al. [23] recognize the difficulties with screening for injury risk as injury patterns are complex and vary over time. Even though screening cannot predict, injury Verhagen et al [23] argue that screening still remain an important and essential part of our efforts to protect an athlete from injury, and points out that we have to continue to improve and develop our methods and understanding for the complexity of injury prevention and the clinical context in which we applicate our questions.”

5. Page 4 lines 10-11: It would be important to include some of the reliability of assessing knee medial to foot during a SLS to show there is some reliability in a subjective assessment and it helps to rationalize why you used this for the construct validity.

We agree with the reviewer that it is important to include some reliability for assessing the SLS and it helps to rationalize why we used the SLS as construct validity. We have therefore made amendments accordingly.

Amendment in background: page 4 line 9-20.
“Some authors propose a simple segmental approach as they only assess the relation between the foot and the knee [24], whilst others propose a multi segmental approach, assessing the whole kinetic chain from the foot to the trunk [25]. A medial displacement of the knee, where the knee during a loaded position moves medial to the foot, is also known as a dynamic knee valgus [19, 26, 27] and is characterised by excessive pelvic drop, femoral internal rotation, knee valgus, tibia internal rotation and foot pronation [26, 27]. Visual assessment of the knee in relation to the foot have been found to be reliable and valid in a clinical and research setting for an asymptomatic adult population [20, 21, 24, 28] whilst the multi segmental approach has been questioned [20, 21]. Recently, the multi segmental approach, preferably with a ≤3 point rating scale, has been found reliable [28]. The aim of an assessment of more than one segment is to give the clinician further information which might be useful in the targeted rehabilitation [21, 28].”

6. Page 6 Line 4: Is the best discriminative ability of the Qinematic software to discriminate between knee over foot and otherwise? Please include this as the discriminative criteria in this line.

This might be a slight misunderstanding due to ambiguity from our side, the best discriminative ability of QinematicTM might not be to discriminate between knee and foot or inversely. The purpose of our study is to explore which angle or cut off point, measured by QinematicTM, that in the best way can match (validate) a visually assessed SLS. Due to this, a clarification has been stated in the aim.

Amendment in background: page 7 line 11-14.
“A further aim was to identify different angles or cut-off points of medial knee displacement, during a SLS measured by QinematicTM, that in the best way would match the results of a visually assessed knee over foot or knee medial to foot position.”
Page 6 line 9: Why did the authors choose to evaluate the SLS 6 to 7 days between sessions?

This is an interesting question that unfortunately doesn’t have any precise answer. There are not too many papers that have argued for the best time interval between a test-retest situation. Streiner et al. [33] recommends 2-14 days between a test retest situation. In the best of both worlds, the retest should be done not too soon after the first test, in order to avoid spill-over effects of learning, muscle fatigue, delayed-onset muscle soreness (DOMS), etc. but also not too late, since the external conditions (hormones production, weather, physical and mental status, etc) are supposed to be equal at both test occasions. Further, most reliability studies on the Single Leg Squat (SLS) with a test retest situation were executed with a time interval of 1-2 weeks [32]. The performance of a SLS are probably quite consistent over time if the test person is not trained or instructed in how to perform the test. So, the time interval could probably have been longer, if we found it reasonable, but not shorter as in the same day. We hope the reviewer are satisfied with this answer and we have added a reference to the choice of our time interval.

Amendment in the method: page 8 line 5.
A reference to Streiner et al. is given “[33]”

Page 6 Line 14: Please add what information on what the software is trying to discriminate similar to the comment on page 6 line 4.

For an explicit explanation of what the software is trying to discriminate, please see item 6. Following clarification have been done in the manuscript.

Amendment in the method: page 8 line 7-9.
“Different angles, or cut-off points, of medial knee displacement from the QinematicTM data were used to compare the visually assessed knee over or medial to foot position.”

Page 6 Line 17: Could a sex mismatched subject pool contribute to the results? Males exhibited smaller ML knee displacement than women in your 5th reference. Would the reliability results change if you removed the males from this study?

We thank the reviewer for this interesting remark. We performed all analyses stratified for sex and we did not find any significant (i.e. meaningful) differences between men and women in the absolute reliability except for the variable left knee down (SEM: men=7.74, women=10.45). All in all, this indicates that in absolute numbers, the group could be analysed together.

For the relative reliability, we found a difference between the sex for one out of four variables, and this was for the variable left knee “down” which showed a very strong Spearman correlation coefficient (r=0.93) in men, but only a fair correlation (r=0.55) in women. For this variable, the ICC was 0.86 in men, while only 0.59 for women. For both men and women there was no significant difference between test occasion one and two for the variable left knee “up” as it was for the whole group men and women. Perhaps this lack of statistical difference is due to a lack of power, since the Bland-Altman plots still indicated systematic differences between the two occasions for the variable left knee “up” for women.
(6.101, 0.441 to 11.761) but not for men (9.264, -3.239 to 21.767). All in all, we believe that there are no reasons to believe that there are any great differences in performance between the sexes and the lack of power restrict us to present the results stratified.

10. 
Page 6 Line 20-21: Does your exclusion criteria also include total knee replacement?

Thank you for pointing this out. Of course, we did not include people with total knee replacement, and agree that this should be stated.

Amendment in the method section: page 8 line 15-16. 
“(ligament- or meniscal rupture and knee replacement)”

11. 
Figure 1: A drawback of the using the Kinetic System is the necessary space and stable background. Not all clinics may afford those setups. This language would be beneficial in the discussion along with the points that the Kinect system does not add anything substantial to a single leg squat assessment that a clinician can observe already.

See item 34 for an answer and an amendment:

12. 
Page 8 line 24: Was there a minimum depth needed to achieve a valid trial?

We thank the reviewer for this interesting note. There are today a great variety of depth in the performance of a SLS and no standard depth mentioned in the literature [32]. QinematicTM has a prespecified depth for its double leg squat (personal communication with the company) but not for the SLS. Following amendments has been done in the manuscript.

Amendment in the method section: page 10 line 23-24 and page 11 line 1. 
“QinematicTM has no prespecified depth of knee flexion to be achieved during the SLS, but the non-weight bearing leg is not allowed to cross the midline.”

13. 
Page 9 Line 11: Did the reviewers watch both the sagittal and coronal plane videos together or separately?

Yes, this is an important remark about the visual assessment and the videos were presented in the same recording.

Amendment in the method section: page 11 line 19. 
“The videos were presented in sagittal and frontal plane in the same recording.”

14. 
Page 9 line 12-14: What were the rating criteria used on how to assess the tests? Could the authors provide a table on these criteria and how the criteria contributed to the construct validity. What were the outcomes the reviewers were assessing on the single leg squat?
This has in previous manuscript been reported in the method section under the heading “variable and data management”. After reading the reviewers comments, we find it more appropriate that this clarification comes earlier in the manuscript. The text from the method section “variable and data management” has therefore been moved, and slightly changed, to the proposed position.

Amendment in the method section: page 11 line 11-18.
“The rating of the SLS was dichotomized (fail/pass), and a subject was scored as either having a knee over foot position (pass), or a knee medial to foot position (fail). A knee over foot position was scored as a pass when the knee was well aligned over, or lateral to, the second toe, and a knee medial to foot position was scored as a fail when the knee was placed medial to the second toe [24]. The visually assessment of the SLS is reported as valid and reliable for use in research and clinical settings for an asymptomatic adult population [20, 21] and might therefore be used as a construct in a construct validity study.”

A small complement has also been done in the method section under the heading “variable and data management”.

Amendment in the method section: page 12 line 20-22.
“The medial displacements, measured in degrees by Qinematic™, was then compared to the dichotomized (fail/pass) visual assessment of the knee position in relation to the foot as described above.”

15. Page 9 Lines 14 -15: Were these videos restricted to the first session or second session or across all sessions?

We thank the reviewer for this remark. A clarification has been done accordingly.

Amendment in the method section: page 11 line 23.
“across all trials (i.e. right and left leg for two test occasions)”

16. Page 9 Lines 15-16: Can the authors clarify if all 11 videos were individually assessed or all videos of all subjects were assessed (i.e. 37 if only 1st session or 74 for both sessions)?
Yes, the same clarification as in item 15 has been done.

Amendment in the method section: page 12 line 1.
“, across all trials,”

17. Page 10 lines 7 -8: Are there any practical reasons why the construct validity analysis was limited to "the way down"?

Thank you for pointing this out.
We limited the construct validity analysis to “the way down” since “the way up” showed poor reliability with a significant difference between the test-occasions. Acceptable reliability is a pre-request for validity, since a test with a good validity but poor reliability cannot be trusted.
A smaller clarification and rewriting have been done.
“The reason for this was that the “way up” showed poor reliability with a significant difference between the test occasions in the reliability study.”

18. Page 10 lines 11-12: Move “fail/pass” immediately after dichotomized.

Thank you for this comment which makes it clearer what is meant with dichotomized.

Amendment in the method section: page 12 line 19.
“dichotomized (fail/pass)”

19. Page 11 lines 5-7: Move these lines to the beginning of the paragraph and state if the data were normally distributed. My assumption is that they weren't based on the median, IQR and Wilcoxon sign-rank test.

See explanation and amendment to this item under item 2.

20. Page 11 Lines 9-10: The reviewer noticed that the authors used medians and IQRs as there descriptive and used Wilcoxon signed-rank tests to test for differences between test sessions 1 and 2. The authors should specify if they had non-normal data for the ICCs. If they had non-normal data were the data transformed and did the transformed data satisfy assumptions for normality? Otherwise it is difficult to interpret the ICCs presented in this study.

See explanation and amendment to this item under item 2.

21. Page 11 Lines 20-21: It would be helpful to include the bland altman plots in a supplementary section for readers to review.

Yes of course, we agree with the reviewer that it would be helpful for the readers if the Bland-Altman plots were included and have therefore added them in the additional file 2.

Amendment in the method section: page 14 line 6.
“Additional file 2.”

Amendment in the manuscript: page 36 line 8-12.
“Additional file 2
File format: .pdf
Title of data: Bland-Altman plots reliability study
Description of data: Contains Bland-Altman plots for the four variables left knee up, left knee down, right knee up and right knee down.”

When calculating new Bland-Altman plots, we found that this have been done in reversed order, i.e. test occasion 1 minus test occasion 2. The positive and negative values of the mean difference and limits of agreement has therefore been changed in table 2 after calculating new Bland-Altman plots in correct order, test occasion 2 minus test occasion 1.
Amendment in table 2: Positive and negative values of the mean difference and limits of agreement has been changed.

Results:
22.
Page 12 Line 24: I am assuming that (#28 are the number of measures that were performed on easy mode on the first of second test session. Consider using a fraction (28/32) for clarity and consider this approach for the paragraph. Were any measures tested on easy mode during the second test session and how were those data handled?

Thank you for your comment and suggestion. We have made changes accordingly.

Amendment in the method section: page 15, line 6-7 and line 13-16.
“Thirty-two measures were excluded due to the tested person performing an “easy mode” on the first or second test occasion (28/32) or an “easy mode” on both occasions (4/32).”

“Missing data were classified due to; poor video quality (4/20), the tested person losing their balance (2/20) and missing video recordings (13/20). For one test person it was impossible to know which SLS was to visually assess, since this subject performed two SLS on the video recordings but only one was recorded by QinematicTM (1/20).”

Unfortunately, we found two mistakes in this paragraph. It regards the available data in the reliability and validity study, the percentage figures were mixed up and written wrong. This have now been corrected.

Amendment in the method section: Page 15, line 10-11 and 16-17.
“Thus, 85% of the data were available for the test-retest reliability study.”
“In total, 76% of the data were available for the validity study.”

23.
Page 13 lines 5-6: Spelling - Numerics should not start a sentence.

Of course, these have now been written out.

24.
Page 13 lines 7-8: Is this indicating that the software did not match a video on one of the subjects? Please clarify.

Thanks for pointing this out, the writing was unclear. A measure was recorded by QinematicTM but the subject performed two SLS on the video recordings which made it impossible to know which SLS to assess.

Amendment in the method section: Page 15 line 14-16.
“For one test person it was impossible to know which SLS was to visually assess, since this subject performed two SLS on the video recordings but only one was recorded by QinematicTM (1/20).”

25.
Table 2: Please clarify the n's in this table. Do they represent the number of subjects or measures?
Thanks for pointing this out. This has now been clarified

Amendment in table 2:
“n=denotes the number of measurements done by QinematicTM for each variable”

26.
Page 15 Line 2: Could the authors please indicate what positive and negative values indicate (are they medial and lateral deviations from the start position?)

We thank the reviewer for this comment as this was not obvious at all. Following clarification have been done.

Amendment in table 2:
“A negative value (-) denotes a medial displacement for the left knee and a lateral displacement for the right knee, contrariwise a positive value (+) denotes a medial displacement of the right knee and a lateral displacement for the left knee.”

27.
Page 15 Lines 1-10: This information is redundant from the table. Please remove and consider referencing table 2 in the next paragraph.

We thank the reviewer for this comment and have now removed the paragraph in full. The referencing of table 2 and the second paragraph are now before the table as we changed the position of those two.

Amendment in the results section: page 15 line 20.
“All psychometrics from the test-retest reliability study are presented in Table 2.”

28.
Page 15 lines 13-14: Consider adding that these variables did not differ statistically between the two occasions and include the lowest p value.

Thank you for this remark. We have corrected the paragraph.

Amendment in the results section: page 15 line 25.
“did not differ statistically between the two occasions (p=0.21).”

29.
Page 15 lines 18-20: Adding to my previous statement on page 11 supplying the Bland Altman plots as a supplementary file.

In line with item 21 the Bland-Altman plots are placed in additional file 2.

30.
Page 16 lines 9 & 10: A similar comment to the table, why is a 15 and 17-degree cutoff being reported when cutoffs were applied over 20 degrees in 2 degree steps?

We thank the reviewer for this remark, this was an illogical description of the cut-off points from our side. The cut-offs 15° and 17° have now been removed from table 3 and replaced
with the cut-offs 14°, 16° and 18°. Due to these changes in table 3 additional changes have been done in the result section.

Amendment in the results section: page 18 line 10-12.
“0.58 (14° cut-off)”
“86% (16° cut-off)”

Discussion:

Page 17 Lines 16-18: Could the authors include some of those reliability results from the other studies with the caveat being differences in how the kinematics are calculated. It would help to frame the current studies results and provide some evidence if there it is a hardware or software limitation.

We thank the reviewer for this comment and agree that it is interesting to put the method that the QinematicTM system is using in the light of previous studies. We found five studies that reported validity [29-33] and test-retest reliability [31-33] for a SLS or a double leg squat (DLS).

Starting off with the validity part in these studies, all studies except for Wochatz et al. [33] compared the Kinect V2 Software Development Kit (SDK) together with a specific customized post-processing technique against a laboratory-based advanced video-analysed system. The major differences between the abovementioned studies and the present study on QinematicTM is thus the post-processing method of measuring lower extremity kinematics. More specific, the previous studies searched for one peak joint angle while the QinematicTM system is using all angles of the movement. QinematicTM calculates the Net Trajectory Angle (NTA), which estimates the “line of best fit”, and is an average of medial and lateral displacements measured 30 times per seconds from the top of the squat to the bottom of the squat and vice versa after the turning point. This use of post-processing (software issue) has of course a large impact on the validity (what are you measuring), but also on the stability of the measurement (reliability). Another reason for poor validity against laboratory-based advanced video-analysed systems is that transforming knee angles to a 2D-plane (the frontal plane) is found to be very difficult when both femur and tibia are rotating in- and/or externally during the movement and Mentiplay et al. [31] found an ICC of 0.02 for knee valgus during a SLL, indicating that such knee valgus is difficult to be captured by the Kinect camera. This argues that there are both a hardware and a software issues involved in problems with validity.

Concerning the reliability part, we believe that the high values of SDC showed by QinematicTM (i.e. poor reliability) could thus also be the result of the poor validity of the NTA. Very small variations in medial/lateral knee displacement during the SLS, could result in a large NTA changes due to the “line of best fit” during the whole movement. When only measuring one knee abduction angle there could be a relatively small within subject- and between days kinematic variation [21], but by measuring the NTA it seems that QinematicTM captures all small variations during the whole movement, resulting in low reliability. Interestingly, Wochatz et al. [33] found that smaller movements had lower reliability, since the measurement error is proportional larger in small movements compared to large movements. Again, these findings argue for the existence of both hardware and software issues when assessing the QinematicTM reliability.
Amendments in discussion section: page 19 line 18-24 and page 20 1-6.

“This due to all studies, except for Wochatz et al. [46], compared the Kinect V2 Software Development Kit (SDK) together with a specific customized post-processing technique against a 3D laboratory-based video-analysis system. In other words, the major differences between those studies and QinematicTM is thus the post-processing method of measuring lower extremity kinematics where previous studies search for one peak joint angle while the QinematicTM system uses all angles of the movement. This in order to calculate the net trajectory angle (NTA) which estimates the “line of best fit”, and is an average of medial and lateral displacements, measured 30 times per seconds from the top of the squat to the bottom of the squat and vice versa after the turning point. In terms of absolute reliability Mentiplay et al. [43] measures the knee abduction angle of a SLS (SEM=4.38°) and vertical drop jump (SEM=3.62), Schmitz et al. [44] measure knee adduction (MDC=3.1°) and knee flexion (MDC=4.1°) in a double leg squat and Wochatz et al. [46] measure knee flexion/extension (SEM=6.8-8.3°) in a double leg squat.”

Amendment in the discussion section: page 22 line 2-8.

“In terms of reliability, the above-mentioned studies highlight difficulties with both the hardware and the software parts of the QinematicTM system during measuring knee angles in the frontal plane during concurrent internal and external rotations of femur and tibia during a SLS. Moreover, the post-processing algorithms of QinematicTM (the NTA) which enables to capture the whole movement in one point, seems to be more unreliable compared to attempts of capturing peak angles at one specific point during the movement.”

32.
Page 18 Lines 8-9: Can the authors provide some examples of a simple and complex movements. My assumption is a single leg squat is a complex movement that the Kinect struggles with but what is a simple movement?

We thank the reviewer for this comment. This information was referred from the study of Wochatz et al. [34] who showed that the variability of their results increased with the complexity of the movement task in following order: Double Leg Squat-hip abduction-lunge. How complex a Double Leg Squat (DLS) is in relation to a hip abduction could of course be discussed but the DLS would probably be a simpler movement than a lunge.

Amendment in the discussion section: page 20 line 17-18.

“(that is order: Double Leg Squat-hip abduction-lunge)”

33.
Page 19 Lines 8-9: This helps to address my previous comment in the methods about depth, but as you state it is clearly a drawback on your current study.

Yes, we agree with the reviewer, we have also made an amendment about the depth of the SLS in the method section (see item 12).

34.
Page 22 Lines 7 - 8: The reviewer agrees with the lack of a 3D kinematic gold standard as important limitation in the current study and agrees that the construct validity assessment is the next logical choice. The discussion is missing a section on what the addition of this software brings to the clinic. If all the software can be good for is to assess that someone does not have their knee go over their foot during a SLS then there is no need for this software in
the clinic. A rater can visually inspect if the knee travels medial to the foot during a SLS descent. The software does not provide reliable and accurate kinematic data that may be needed for more refined kinematic analysis of injury risk. This language should appear more in the discussion.

We thank the reviewer for this important remark about the clinical implication of QinematicTM and we agree that the software, for the moment, doesn’t add anything substantial to the clinical situation. Amendments have been done accordingly.

Amendment in the discussion section: page 24 line 3-9.
“From a clinical perspective one could argue that QinematicTM, does not contribute to anything additional as a clinician is able to visually assess a knee over foot position with a good accuracy. Thus, another kinematic analysis would have been desirable in order to add something substantial to the clinical setting. As suggested by Eltoukhy et al. [29], perhaps the most valuable capacity of QinematicTM of today, in the sense of knee assessment or knee rehabilitation, would be as a pedagogic tool where the patient could see the trajectory of the knee as it moves medial or lateral.”

Conclusions:
35.
Page 22 Lines 15-16: consider re-wording to “reliability when measuring a SLS on the way down not at the way down.

Thank you for your comment and suggestion, this have now been done.

Amendment in the discussion section: page 25 line 15.
“not at the way up.”

36.
Page 22 line 25: Could the authors include examples of additional tests, like the vertical drop test they mentioned earlier, that can be used to assess knee kinematic deficits.

Thank you for your comment, this suggestion is now corrected.

Amendment in the discussion section: page 25 line 24 and page 26 line 1.
“perhaps a vertical drop jump [35] or similar test that places a higher demand on the knee.”

37.
Table 3: The methods described that pass/fails were quantified using 2-degree steps from 2 to 20 degrees. Why are the 14, 16, and 18-degree steps not included in this table and the 15 and 17 degree steps used?

See item 30 for an answer and an amendment:

38.
Figure 2: Check spelling in figure legend.

Yes, this has been done.


