Peer review file

Article information: https://dx.doi.org/10.21037/jss-21-16

Reviewer A

Comment 1: Line 128. Please put up the literature on Spinal-stenosis related gait alterations (SSRGAs).

Reply 1: Yes, we should cover literature related to LSS-related gait alterations in the Introduction. Note that we have changed the scope of our manuscript to focus only on lumbar stenosis instead of spinal stenosis more broadly.

Changes in the text: We have added on pages 4-5, lines 95-98 in the Introduction, which broadly covers the LSS-related gait changes that have been reported by other studies in the literature. Our Discussion on pages 10-11, lines 225-240 explores these findings in more detail.

Comment 2: The sample size may have been smaller than planned due to COVID-19, but is it reasonable to lump together gait disorders due to cervical spinal stenosis and gait disorders due to lumbar spinal stenosis as SSRGAs? Central stenosis of the cervical spinal canal causes central nervous system dysfunction, while lumbar spinal canal stenosis causes peripheral nervous system dysfunction, thus the pathology of the two is different. The appropriateness of lumping them together needs to be explained.

Reply 2: Yes, we agree that the pathology of CSS and LSS are different and lumping the two conditions together is inappropriate.

Changes in the text: We have decided to remove analysis related to CSS from this paper, which we plan to use for a future separate paper. Our paper is now focussed only on LSS, rather than spinal stenosis in general.

Comment 3: Lines 130-133. Relating to the above point, references 37, 44, and 49 are reports on patients with lumbar spinal stenosis and the older adults. Please also cite the report on patients with cervical spinal stenosis.

Reply 3: As above, we have decided to remove CSS from the scope of this paper.

Changes in the text: Our paper is now focussed only on LSS, rather than spinal stenosis in general.

Comment 4: Lines 187-201. I am not an engineering specialist, so I cannot comment in detail about wearable devices and analysis software, but please tell me how spatial parameters (stride length) can be obtained from an IMU attached to the chest.

Reply 4: Gait analysis involving spatial parameters such as stride length has been previously performed using chest-based sensors, such as in the study by Hashmi, Riaz, Zeehsan, Shahzad & Fraz (reference 59 of our revised manuscript). In addition, a systematic review (reference 57 of our revised manuscript) investigating the measurement of gait in Parkinson's disease found that 9 papers used an IMU system involving chest-based sensors to measure gait parameters such as the ones used in our study. In a similar manner, data from our MMC sensor (obtained using an accelerometer for the measurement of linear acceleration, a gyroscope for the measurement of angular acceleration, and magnetometer for the measurement of the orientation of the Earth's gravitational field) can be paired with software analysis to produce spatial and temporal gait parameters. Our accuracy analysis has partially validated our approach, although future studies are required to assess the accuracy of our IMU system for the measurement of variability and asymmetry (as implicated in page 15, lines 344-347).

Changes in the text: We have added on page 10, lines 348-374, a paragraph to justify our use of a chest-based IMU.

Comment 5: Lines 213-217. Video recording is used as the gold standard in this study. Please describe in more detail the specs of the video camera, the method of recording, and also how you measured the number of steps and walking time.

Reply 5: Yes, we should have provided more detail regarding the methodology behind our accuracy analysis.

Changes in the text: We modified the text as advised in pages 6-7, lines 146-158 in our Methods.

Comment 6: Lines 244-246. Please describe in detail your process in determining the cutoff value for the WORM score. For example, the results of the ROC analysis.

Reply 6: Yes, more detail was required. However, we feel that providing additional detail regarding the WORM score in this paper would decrease the overall clarity of this paper. Therefore, we intend to discuss the WORM score in a future paper.

Changes in the text: We have removed the WORM score from our paper.

Comment 7: Lines 254-264. Please indicate the radiological severity of the subject so that you can more clearly characterize your cohort.

Reply 7: Yes, providing the radiological severity of each subject would enable us to better characterize our study cohort. However, the current study is only a preliminary analysis of the gait of LSS patients with a relatively small sample size that does not lend itself to sub-analysis based on radiological severity. We agree that this is a limitation of our study, since LSS patients with varying levels of radiological severity would produce varying gait patterns.

We are currently performing a follow-up study with a larger sample size and with sub-analyses based on radiological severity.

Changes in the text: We have now identified this as a limitation of our study on page 12, lines 274-276. We have highlighted the need for follow-up studies with larger study populations to analyse gait patterns amongst the radiological subgroups of LSS on page 12, lines 277-280.

Comment 8: Table 3. I think it would be useful to show graphs of the Bland-Altman plot. Even though robust statistical analysis is not possible, it would be useful to describe the measurements by subtypes (cervical central canal stenosis, lumbar central canal stenosis, lumbar lateral recess/foraminal stenosis).

Reply 8: Yes, it is useful to show the Bland-Altman plots and it is important that gait measurements by subtype are quantified. With our small sample size, we believe that these values by subtype may be misleading. We are currently performing a follow-up study with a larger sample size which would enable us to quantify LSS gait changes by subtype.

Changes in the text: We have added all Bland-Altman plots to the Supplementary Appendix. Page 12, lines 277-282 also indicates the need to perform a follow-up study with a sufficiently large sample size to quantify gait changes by LSS subtype.

Reviewer B

Major

Comment 1: To test the accuracy of a novel system necessitates a valid, accurate and reliable gold standard. The authors compared spatiotemporal gait parameters assessed using an IMU system with videography. Although the authors cite literature that uses camera based systems, those studies used markers and cameras to determine gait parameters. In the present study, the authors used basic video footage to determine these parameters. Not only is the methodology of obtaining spatiotemporal parameters using this method insufficiently described in the manuscript, the validity accuracy and reliability of this mistakenly called gold standard system is unknown. Moreover, the authors do not specify how the outcomes are defined and computed using this system or how data for multiple steps in a walking trials were postprocessed and treated in the statistical analyses. If the authors wish, they could call their comparison a method comparison but the study aim of testing the accuracy of a chest-based wearable sensor, the MetaMotionC© device and inertial measurement unit python script (MMC/IMUPY) system can not be reached in the current study design. Finally, the authors did not address their aim of testing the reliability of a chest-based wearable sensor, the MetaMotionC© device and inertial measurement unit python script (MMC/IMUPY) system.

Reply 1: Yes, we were incorrect in describing our videography technique as the gold-standard. We also have not addressed our aim of testing the reliability of the MMC device.

We should also provide more information on the MMC system and its related data processing steps.

Changes in the text: We have removed our aim of testing the reliability of the MMC device from our revised manuscript. We have also now termed our use of videography to be a reference-standard instead of a gold-standard. We have included the specifications of our videography technique (on page 7, lines 146-159) for additional detail. We have also reworded our manuscript to state that we are only trying to preliminarily test the accuracy of the MMC/IMUPY system. Pertaining to this, we have also acknowledged (on page 15, lines 344-347) that true gold-standard gait analysis techniques such as laboratory systems and gait mats should be used in future studies to assess the accuracy of the MMC system. Appendix 1 in our attached Supplementary Appendix document provides more information regarding the MMC system and its related data processing steps. We also flagged this in page 6, line 133-134.

Comment 2: The authors included a rather small and very heterogeneous cohort. Patients with spinal stenosis included patients with cervical and lumbar stenosis who could show distinctly different gait patterns. The authors did not report the severity of the stenosis, which presumably impacts spatiotemporal gait parameters. The control group comprised patients who had received treatment but did no longer have symptoms. It is unknown if these persons return to normal gait patterns. While the choice of control group was convenient for the authors as these persons visited their clinic anyways, a truly asymptomatic control group should have been included that most importantly also matches the age of the patient group.

Reply 2: Yes, cervical and lumbar stenosis patients could show distinctly different gait patterns. Yes, a truly asymptomatic control group should be included. While the radiological severity of the stenosis likely impacts gait parameters, we did not report mean gait parameters by subtype of radiological severity as we believe that, with this small sample size, the results would be misleading. Instead, we are currently performing a larger follow-up study where we will be able to analyse gait patterns in LSS patients by subtype of radiological severity.

Changes in the text: We have removed cervical stenosis from the scope of this paper and intend on performing a future study dedicated to gait alterations in patients with cervical myelopathy. We have recruited additional participants via verbal outreach to include 25 truly asymptomatic healthy controls with similar ages to the 25 LSS patients. All sections in Results have been updated and statistical analyses have been redone.

Comment 3: The authors correctly state in their limitations that age affects gait patterns. However, this known effect must be considered and addressed in a study design for a study aiming at identifying the changes to gait patterns that occur in spinal stenosis. (The authors did not quantify changes in gait patterns with the disease, which would require a longitudinal study design, but differences in gait pattern between older patients with symptomatic spinal stenosis of varying location and unknown severity with younger patients after successful

treatment for symptomatic spinal stenosis of varying location and unknown severity. Calling the control group ‘healthy’ is misleading.

Reply 3: Yes, our study participants in the control group were unsuitable. As above, we have recruited 25 new truly asymptomatic healthy controls via verbal outreach with similar ages to our LSS cohort.

Changes in the text: All sections in Results have been updated and statistical analyses have been redone.

Comment 4: The study lacks a sample size estimation. Data of several participants had to be excluded because of insufficient data. The authors justify the size of their cohort by the difficulty of recruiting participants because of the Covid pandemic. Scientists around the world are challenged by delays in patient recruitment. However, this does not justify presentation of data of underpowered studies.

Reply 4: Yes, the sample size of our LSS cohort is a limitation of our study. We intend for this to be a preliminary study for the construction of an early DSGP for LSS. We are currently performing a larger follow-up study to provide a more robust DSGP for LSS.

Changes in the text: None.

Comment 5: The authors report the results of Bland-Altman plots in Tables 3a and 3b. However, Bland-Altman also contains information on errors depending on magnitude of parameters that are not presented in the Tables. This information should be provided.

Reply: Yes, we omitted some information from the Bland-Altman plots in our Tables.

Changes in the text: We have now included the complete Bland-Altman plots in Appendix 2 of our Supplementary Appendix.

Comment 6: The authors determined the cut-off of the WORM score based on retrospective data on falls in this specific cohort. Based on the limitations stated above and the retrospective nature of fall risk, the generalizability of this cut-off is largely limited.

Reply 6: Yes, our analysis regarding the WORM score was very limited. We should be more robust in our justification of the WORM score and derivations of falls risk stratification, which we have decided to relocate to a separate publication.

Changes in the text: We have removed the WORM score from this paper.

Minor

Comment 1: The authors should clearly state that only spatiotemporal gait parameters are assessed in their study.

Reply 1: Yes, we should make it clearer that we are focussing on only spatiotemporal gait parameters.

Changes in the text: We have modified the title of the revised manuscript “Objectifying clinical gait assessment: Using a wearable device to quantify the spatiotemporal gait metrics of people with lumbar spinal stenosis”.

Comment 2: Introduction: The rationale for using IMUs for assessing gait in this specific population is missing.

Reply 2: We have now made this clearer.

Changes in the text: We have made more succinct our explanation of the rationale for using IMUs to assess gait in this specific population on page 3, lines 65-78.

Comment 3: Clear inclusion and exclusion criteria should be stated.

Reply 3: While we mentioned our inclusion and exclusion criteria in our Methods, we agree that this could have been done more clearly.

Changes in the text: We have added a new Table 1 which clearly details the inclusion and exclusion criteria for our LSS cohort.

Comment 4: Lines 177-178: “…or IMUGait application bugs (19 occurrences)…” Please provide concise criteria for data quality.

Reply 4: In these instances, we excluded patients due to a total loss of data from the walking bout when the IMUGaitPY application malfunctioned and failed to download gait data from the sensor. These data were unrecoverable. It was not due to any lack of quality of the data obtained.

Changes in the text: Clarification in page 5, line 113 that we were referring to a loss of data.

Comment 5: The sensor uses information of a magnetometer. A well-known problem of using magnetometer data is the sensitivity to the ferrous materials the environment and resulting decrease in accuracy. Was the hallway tested for such interference?

Reply 5: The environment was not tested for ferric interference.

Changes in the text: We have now acknowledged this as a limitation of our study and have added this consideration to our Discussion in page 14-15, lines 338-343, and page 15, line 347.

Comment 6: Please provide sufficient detail on the videography (hardware, software, computation of outcomes).

Reply 6: Yes, it would be better to add detail on the videography method used.

Changes in the text: We have modified the text in our Methods as advised on page 4, lines 124-126.

Comment 7: The WORM score was inspired by several papers which used the motion of the centre-of-mass during quiet stance (standing still). The IMU was placed on the sternum which is not the COM. Please provide rationale and justification for placing the sensor at this location.

Reply 7: We have removed the WORM score from this paper. Besides this, we believe that the chest-based placement of our sensor is particularly viable for future possibilities involving patient-use in everyday living conditions.

Changes in the text: We have added a paragraph in our Discussion on pages 15-16 lines 348-374 discussing the use of chest-based sensor placement.

Comment 8: Please stratify the data for age.

Reply 8: Due to this being a preliminary study and our limited sample size, we believe that stratifying values by age bracket will be misleading. However, our significantly different mean age between controls and LSS patients is an issue which we have now addressed by recruiting additional controls. We are also currently performing a larger study where age stratification will demonstrate meaningful results.

Changes in the text: We have recruited additional participants via verbal outreach to include 25 truly asymptomatic healthy controls with similar ages to the 25 LSS patients. All sections in Results have been updated and statistical analyses have been redone accordingly.

Comment 9: Bland-Altman plots should be presented.

Reply 9: Yes, we omitted the Bland-Altman plots.

Changes in the text: We have now included the complete Bland-Altman plots in Appendix 2 of our Supplementary Appendix.

Comment 10: Did walking speed affect the agreement between systems?

Reply 10: It does not appear that walking speed affected the agreement between the two systems. We did not investigate this further with statistical methods.

Changes in the text: None.

Comment 11: Agreement between two poor methods does not proof accuracy of either system. Please rephrase the manuscript to reflect that two methods were compared and ensure that the strength of the conclusions match the results of the study.

Reply 11: Yes, we now realise that our conclusions were reaching.

Changes in the text: Throughout the text, we have now referred to videography as our reference-standard, and not gold-standard. We have also modified our conclusions to state that our sensor had high agreement with videography for all metrics that both methods could estimate, rather than claiming that our sensor has been validated. Additionally, instead of claiming that our methods are suitable for clinical contexts, we have changed our argument to state that our methods have potential for use in clinical contexts. We have also outlined further necessary steps in our Discussion that should be taken before clinical application (such as on pages 12-13, lines 277-295 and on pages 16-17, lines 375-401).

Comment 12: Lines 318: The authors state that “However, clinicians should still consider the adjusted estimation of non-temporal gait features by the MMC/IMUPY to be within 10% of the true value on account of the inconsistency when estimating these metrics.” Please place this result in a clinical context. Would this consistency be sufficient in a clinical context?

Reply 12: We agree that this degree of consistency is not sufficient to detect minor gait alterations. In a clinical context, this prevents our gait analysis system from being able to detect early signs of gait-altering disease.

Changes in the text: We have added this consideration to our Discussion on page 10, lines 286-287.

Comment 13: All discussion points on relating observed differences to symptomatic spinal stenosis should be rephrased because they could be due to the difference in age between groups and also influenced by severity, which was not reported.

Reply 13: Yes, our discussion points could have been affected by differences in age groups and LSS severity. Besides the changes in the text outlined below, we are currently performing a larger study to construct a more nuanced gait profile for LSS, adjusted for LSS severity, age, and sex.

Changes in the text: We have now recruited additional controls with similar ages to our LSS cohort. In addition, we have reworded our manuscript and now only proposing a preliminary gait profile for LSS.

Comment 14: Line 402: "…This is also the first study to quantify the impact of spinal stenosis on postural instability." How was postural stability defined and measured in the present study?

Reply 14: Postural stability in this study related to the WORM score, which we have now removed from the study.

Changes in the text: The WORM score has been removed.

Comment 15: Line 454: "Assuming the WORM score performs equally well in elderly populations, it would rank amongst the top 5% of classification tests." This statement lacks support by data in this manuscript and should be deleted. The authors could have included fall risk questionnaires to assess the value of WORM score.

Reply 15: We have now removed the WORM score from this paper.

Changes in the text: The WORM score has been removed.

Comment 16: The authors discuss the literature of other IMU systems. However, this should be discussed in the context of sensor placement, parameters and study cohorts.

Reply 16: Yes, we should add a discussion of sensor placement and related parameters. We did include a discussion of the utility of IMU systems in different study cohorts in our Discussion on pages 13-14, lines 312-328.

Changes in the text: We added a section in our Discussion on chest-based sensor placement and the gait parameters that have been shown to be captured through this method in the literature on pages 15-16, lines 354-368.

Comment 17: The conclusions are not supported by the data presented in this manuscript and must be revised.

Reply 17: Yes, our conclusions were not in line with our findings.

Changes in the text: We have modified our conclusions such that we are only providing a preliminary disease-specific gait profile which needs to be built upon by larger studies which are able to stratify data by LSS severity, and by age and sex. Also, instead of recommending our system for clinical use, we have outlined in our Discussion additional steps that should be taken before clinical application (such as on pages 12-13, lines 277-295 and on pages 16-17, lines 375-401).

Comment 18: The authors cited 104 studies. Please limit references to those that are necessary for the context of the study. Some important references are missing (i.e. by Haddas et al. https://pubmed.ncbi.nlm.nih.gov/?term=haddas+cervical+gait).

Reply 18: Yes, we should make our manuscript more succinct.

Changes in the text: We have removed cervical stenosis and the WORM score from this paper. We have also cut down on unnecessary references such that there are now only 63 references.

Comment 19: Table 2: Demographics should have been compared between both groups.

Reply 19: We compared demographics between healthy controls and LSS patients and no significant differences were found.

Changes in the text: None.

Reviewer C

Comment 1: The introduction is lengthy - I suggest condensing to highlight key points.

Reply 1: Yes, the introduction was lengthy. It should be condensed to be made clearer.

Changes in the text: We have significantly shortened the Introduction to 678 words and 3 key paragraphs.

Comment 2: The introduction does not focus on the rationale for conducting this study in spinal stenosis. I suggest re-structuring to make it clear why the focus on stenosis and outline why gait assessment is critical in this population.

Reply 2: Yes, our Introduction should be made clearer. Of relevance is that we have now removed cervical stenosis from the scope of this paper as we believe it warrants a study of its own.

Changes in the text: We have modified the text as advised. We have clarified why the focus of this paper is on lumbar spinal stenosis on pages 4-5 lines 87-100.

Comment 3: What is the focus of this paper? I can see three major objectives including validation of a wearable sensor for assessment of gait, development of a DSGP for stenosis and development/validation of the WORM score for falling. Because there are so many objectives, the depth of analysis for each is not sufficient. This also makes the paper very long. I would suggest pulling out the WORM development and analysis as a separate manuscript. This would allow the authors to focus on the validation of the device and identification of preliminary DSGP for stenosis.

Reply 3: Yes, our paper as it was unclear and confusing with so many objectives.

Changes in the text: We have removed both the WORM score and cervical stenosis from the scope of this paper. We believe both warrant their own paper, and their inclusion in this paper made the manuscript convoluted.

Comment 4: I am concerned about the inclusion criteria and population. By including all types of stenosis it makes it difficult to conclude on a phenotype for gait because they all present differently.

Reply 4: Yes, grouping cervical and lumbar stenosis into one analysis was inappropriate.

Changes in the text: We have removed cervical stenosis from the scope of this paper. We have rewritten our Results section and redone all statistical analyses accordingly.

Comment 5: You have included cervical stenosis which does not present with gait limitations. Please explain.

Reply 5: Yes, our inclusion of cervical stenosis was inappropriate.

Changes in the text: As above, we have removed cervical stenosis from the scope of this paper. We have rewritten our Results section and redone all statistical analyses accordingly.

Comment 6: Central stenosis and lateral stenosis are distinct in their gait patterns. Why did you include both? Did you control for the gait patterns based on type of stenosis?

Reply 6: We did not control for gait patterns based on type of lumbar stenosis. We intend to construct only a preliminary gait profile (and have reworded our aims and conclusions accordingly) for lumbar spinal stenosis. We are also currently performing a larger study, where we will be able to create a more nuanced gait pattern for lumbar spinal stenosis, adjusting for LSS severity, the type of lumbar stenosis, as well as age and sex.

Changes in the text: We have reworded our aims and conclusions to state that we are constructing only a preliminary gait profile for lumbar spinal stenosis.

Comment 7: How did you identify those with gait abnormalities

Reply 7: Our eligibility criteria were not clear.

Changes in the text: We have clarified our eligibility criteria in Table 1, detailing how we selected our LSS patient cohort. We have also recruited additional controls via verbal outreach so that all of our controls are asymptomatic individuals from the community.

Comment 8: What were your specific inclusion criteria

Reply 8: Our eligibility criteria were not clear.

Changes in the text: As above, we have clarified our eligibility criteria in Table 1, detailing how we selected our LSS patient cohort. We have also recruited additional controls via verbal outreach so that all of our controls are asymptomatic individuals from the community.

Comment 9: Is this degenerative stenosis? The age seems low. The age of healthy controls is very low (38) and raises concern when comparing with stenosis. Of course there will be differences between older adults with stenosis and 38 year old healthy controls. The results would be more compelling with age matched controls.

Reply 9: We did not restrict our lumbar stenosis cohort to degenerative stenosis. Our inclusion criteria were not clear. We have also recruited additional controls with similar ages to that of our lumbar stenosis cohort.

Changes in the text: We have clarified our eligibility criteria in Table 1. In addition, after recruiting controls with similar ages to our lumbar stenosis cohort, we have rewritten our Results section and redone all statistical analyses.

Comment 10: Why did you choose the Meta Motion device?

Reply 10: The MetaMotionC device was able to measure a suitable battery of spatiotemporal gait metrics whilst also being small, lightweight, and cheap. There is a limited amount of literature exploring its accuracy, which is why we compared the sensor with videography in an attempt to provide preliminary validation.

Changes in the text: None.

Comment 11: Has this device been shown to be valid in other gait limited populations?

Reply 11: Not to our knowledge.

Changes in the text: None.

Comment 12: Why the sternal location (vs. hip or back etc.)

Reply 12: Placement at the chest poses some advantages over other positions. Placement of IMUs at the chest have the advantage of detecting less “noise”, compared to ankle- or wrist-based wearables where sensors frequently change orientation due to the degree of freedom of the upper and lower extremities. Furthermore, the flat bony surface of the sternum enables repeatable sensor attachment by even unskilled users, which is important for future possibilities of being used by patients in everyday living conditions.

Changes in the text: We have added some discussion about the use of chest-based sensors on pages 15-16, lines 348-374.

Comment 13: Is Video truly the gold standard here? What about gait mats/pressure sensors?

Reply 13: Good point, videography is not the gold-standard – gait mats and pressure sensors would be the gold-standard.

Changes in the text: We have changed our manuscript and now refer to videography as our reference standard. In addition, instead of claiming to have validated the MMC sensor, we have now changed our conclusions to state that we showed good agreement between the sensor and videography, indicating reasonable accuracy. We have also mentioned in our Discussion on page 15, lines 344-347 that future studies should endeavour to test the accuracy of the MMC/IMUPY system using gold-standard methods such as 3D motion capture or gait mats.

Comment 14: 5m is not a sufficient distance to validate these parameters. Please comment.

Reply 14: Yes, in our study, we only included walking bouts in the accuracy analysis if they were longer than 30m, as approximations made by videography would have a greater margin of uncertainty for shorter walking bouts.

Changes in the text: None.

Comment 15: 30m is too far for realistic testing in a home or clinical environment. If the goal is to have patients do this at home, they will not be able to complete the testing unless outside. This presents issues for patients with balance issues, particularly in cold climates. Please comment.

Reply 15: Good points. Our eventual goal is to have patients wear this at home and undergo 24-hour gait and walking analysis, so we believed that, over the course of a day, most patients would surely walk more than 30m. However, it is true that it may be less common for patients (particularly patients with severe LSS) to walk more than 30m at a time in one stretch, and your point about a 30m continuous walking bout being feasible only if outside is true. We have now acknowledged this as a limitation of our study and have added this point to our Discussion.

Changes in the text: We have added this consideration to our study on page 14, lines 331-336 and on pages 16-17, lines 386-390.

Comment 16: Why was test-retest reliability not examined

Reply 16: Patients underwent walking bouts for as far as they were willing, with a maximum distance of 120m. If the patient decided to stop beforehand, we did not pressure them to undergo subsequent walking bouts enduring pain. If the patient reached 120, we considered this far enough for the accuracy analysis. Still, you raise a good point, and this is a limitation of our study.

Changes in the text: We have acknowledged this as a limitation of our study and have recommended future studies aiming to validate this system to use test-retest reliability on page 15, lines 343-344.

Comment 17: I am concerned that other factors related to gait were not controlled for

Reply 17: Due to our limited sample size, we did not perform regression analyse controlling for other factors related to gait. Nonetheless, this study only aims to construct a preliminary gait profile for lumbar stenosis, and we are currently performing a larger study in which we will be able to perform regression analysis.

Changes in the text: None.

Comment 18: How did you define the most affected limb?

Reply 18: This was determined based on a self-reported questionnaire that was administered to each patient before their walking bout. We are currently performing a larger study where we can construct a more nuanced gait profile for lumbar stenosis, adjusting for factors such as the affected side.

Changes in the text: None.

Comment 19: There is not a clearly defined DSGP. It would be helpful if the authors provided a table with the clearly defined phenotype based on these results.

Reply 19: Yes, our DSGP was not clearly defined.

Changes in the text: We have adjusted the manuscript accordingly by modifying our figures; now, Figure 4 incorporates a table to clarify our preliminary phenotype of LSS based on these results.

Comment 20: You indicate that there is poor reliability for spatial parameters. What is the implication of this.

Reply 20: Yes, our statement that “However, clinicians should still consider the adjusted estimation of non-temporal gait features by the MMC/IMUPY to be within 10% of the true value on account of the inconsistency when estimating these metrics” should have been analysed further with respect to clinical implications.

Changes in the text: We have added the clinical implications of this on page 10 lines 285-287.

Comment 21: Did you not examine asymmetry? This is confusing.

Reply 21: We did examine asymmetry.

Changes in the text: We have modified the manuscript to clarify.

Comment 22: The conclusions are reaching. This is preliminary validation of a device and should not be recommended for use in diagnosis or clinical evaluations until further validated.

Reply 22: Yes, our conclusions were reaching.

Changes in the text: We have modified our manuscript accordingly, now stating that we are only aiming to create a preliminary gait profile for lumbar stenosis. We have also now focussed on laying out plans for clinical application, instead of recommending our system as it currently is for clinical application.

Comment 23: The conclusion suggests that these results indicate that dynamic balance, symmetry and smoothness should be examined in clinical practice. I don't see how these results support this.

Reply 23: Yes, our Conclusion was misleading.

Changes in the text: We have reworded our Conclusion to state that, while promising, further studies are required before our system can be implemented in clinical practice. Our Conclusion no longer suggests that those metrics should be examined in clinical practice. We have also removed the WORM score (and dynamic balance from this paper).

Comment 24: Again, I have concern that you have used a heterogeneous population of stenosis (cervical, central and lateral) making it difficult to support the DSGP you have suggested. I would suggest re-running the analysis with each type separately. As it stands, this does not represent a pattern for a specific type of stenosis making clinical application difficult.

Reply 24: We have removed cervical stenosis from the scope of our paper and have redone the analysis accordingly. We did not rerun the analysis which each type of lumbar stenosis separately as we believe that the resulting gait profiles for each type of lumbar stenosis would be misleading (with our small sample size). For this paper, we decided to present a preliminary gait profile for lumbar stenosis generally. We are currently performing a larger study where we will be able to generate gait profiles by lumbar stenosis subtype.

Changes in the text: We have removed cervical stenosis from this paper. We have reworded our aims and conclusions to state that we are constructing only a preliminary gait profile for lumbar spinal stenosis.

Comment 25: I have concern about the age. Is this degenerative stenosis?

Reply 25: We did not restrict our lumbar stenosis cohort to degenerative stenosis. Our inclusion criteria were not clear. We have also recruited additional controls with similar ages to that of our lumbar stenosis cohort.

Changes in the text: We have clarified our eligibility criteria in Table 1. In addition, after recruiting controls with similar ages to our lumbar stenosis cohort, we have rewritten our Results section and redone all statistical analyses.

Comment 26: Is the control group appropriate for developing a phenotype in this population? The average age of degenerative LSS is 67. This control group age was 38.

Reply 26: Yes, the control group was inappropriate. We have recruited additional controls with similar ages to our LSS cohort.

Changes in the text: We have redone the Results section and the statistical analyses using controls with similar ages to our LSS cohort.

Comment 27: The fact that other factors related to gait were not controlled for is a major issue.

Reply 27: Due to our limited sample size, we did not perform regression analyse controlling for other factors related to gait. Nonetheless, this study only aims to construct a preliminary gait profile for lumbar stenosis, and we are currently performing a larger study in which we will be able to perform regression analysis.

Changes in the text: We have more clearly identified this as a limitation of our study on page 11, line 255.

Comment 28: The clinical implications: What do you suggest based on this data? Should care providers recommend the use of this testing in clinics? Patient homes? As it stands the 30m minimum distance is a concern.

Reply 28: Yes, our detailing of the clinical implications was not clear. We intend for this or a similar system to be eventually used in the home environment for the diagnosis/monitoring of gait-altering conditions.

Changes in the text: We have added a paragraph in the Discussion on pages 16-17, lines 375-401 about how the clinical uses of IMUs such as this system rely on their potential to be

taken home. We have also clarified our Conclusion to state that future steps are required before clinical application in the home environment.

Comment 29: Did you test for feasibility or usability of the technology? If the plan is remote monitoring with these devices this is critical to test with patients.

Reply 29: Yes, it is important to test the usability of this technology for remote monitoring before clinical application. We did not test this in the present study.

Changes in the text: We have detailed the importance of testing the usability of this or a similar system in the home environment with future studies (on page 17, lines 399-401, 411).

Comment 30: The paper does show preliminary validity data for the device in the assessment of gait in stenosis. The authors should focus on this and lay out next steps for further validation and plans to how to integrate into practice.

Reply: Yes, we have now followed this line of thought, as outlined in previous comments.

Comment 31: There are potentially some conclusions that can be drawn regarding a phenotype if analysis is focused by type of stenosis.

Reply 31: Thank you. We have adjusted our manuscript accordingly as per previous comments.

Final remark: Overall this is important work and has the potential to lead to changes in the way we assess and understand function in stenosis. If refined, this paper could form the groundwork for understanding how to assess gait with wearables.

Reply: Thank you for your feedback.