PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Balance deficiencies in women with fibromyalgia assessed using computerized dynamic posturography: a cross-sectional study in Spain
AUTHORS	Perez-de-Heredia, Marta; Huertas-Hoyas, Elisabet; Martínez-Piedrola, Rosa; Palacios-Cena, Domingo; Alegre-Ayala, Jorge; Santamaría Vázquez, Montserrat; Fernández de las Peñas, Cesar

VERSION 1 - REVIEW

REVIEWER	Borja Sañudo Department of Physical Education and Sport. University of Seville. Seville, Spain
REVIEW RETURNED	09-Feb-2017

GENERAL COMMENTS	Thank you for the opportunity to review this paper. The study is oriented to gain insight into the association between functional independence and balance responses in patients with fibromyalgia. My main concern is the sample size. In my opinion, is too small for a cross-sectional observational study. In addition, some points that the authors should consider which may strengthen the manuscript: Abstract 1. Please include the meaning of SOT in the abstract. 2. Please include data of the correlations in the results section. Introduction 1. Overall easy to follow with logic and rationale for fibromyalgia syndrome. However, the rationale does not seem strong enough for the association between functional independence and balance responses in this patients group. 2. In the introduction section you justify the “poor posture” based on pain severity and fatigue and further you discuss the effects on the functional capacity of the upper extremity. This section must be refocus on the aim of the study. 3. If Muto et al. (2015) and Jones et al. (2009) have previously assessed the relationship between postural control and balance self efficacy in women with fibromyalgia, what does this study add? The need for this study must be clearly discussed to the reader. Further, your aims were previously solved in different studies (e.g., Russek and Fulk, 2009), which should be cited. Methods 1. Please justify why the new criteria in the FM were not used.

	2. Depression is a common symptom in FM patients, if you exclude this population you may bias your results. Please report this issue. 3. “Each condition was performed 3 consecutive times.” Please indicate how the results were computed and included for further analysis. 4. Statistical analysis: Did you check for normality? 5. Was the sample size for a Spearman correlation determined using power analysis? Generally a minimum of 29 participants must to be recruited to determine a reasonable high correlation of two variables (to be able to detect effect size to detect correlation coefficient = 0.5 with alpha and power are set at 0.05 and 80% respectively). 6. Discussion 1. The discussion should be focused towards your true findings rather than what others have found previously. While that is important to give context to the findings, the reviewer wasn’t sure of what your findings mean and in which context. Further, witch such low correlations I believe that your statements must be modulated. 2. On page 12 you report that “lower scores were observed in conditions 4 to 6, which are the more challenging conditions, as these correspond with the mobile platform conditions.” But this is also the case in control participants, thus, I´m not sure if your discussion on this topic is suitable. 3. To justify the fact that there is no consensus in the management of these patients you use very old references. In the last couple of years there are many reviews and meta-analysis dealing with this issue. I suggest rewriting this paragraph. The authors should mention the implications for patients’ management if any, but on the view of their results. 4. Again on page 13 authors stated that this is the first study assessing the relationship between balance and functional independence in ADLs in this population group and there are a few previous studies that have not been discussed (e.g., Goes, 2012 or some preliminary results from The al-Ándalus project.). 5. The physiological basis of the deficits observed in these patients must to be discussed in detail. 6. It would be interesting to look at the differences regarding symptom severity. 7. In the limitations authors suggest that the selection of your sample “makes generalization of the results to the general population difficult.” However, they conclude on the general FM population. This must be modulated. 8. How can you suggest multidisciplinary treatments based on your results?

REVIEWER	Shirley S.M. Fong University of Hong, Hong Kong
REVIEW RETURNED	17-Feb-2017

GENERAL COMMENTS	Abstract, Conclusion: ..."resulting in" negative impact... Better use the term "associated with" negative impact. Introduction, Aims of study: to investigate which sensory opponent is the most affected (vestibular, visual or somatosensory)...However, the sensory organization test vestibular, visual and somatosensory ratios were not reported in the manuscript. Must report the 3 sensory ratios. Introduction, Aims of study: With the current results, I think another

	aim of the study was to compare the sensory organization of balance control and balance strategies between women with FM and controls (i.e., the two-way repeated measures group effect). The time and interaction effects were also the aims of the study? Methods, inclusion criteria: instead of stating "NO pain" and "NO history" (move these to the exclusion criteria), should include age range and other demographics such as body mass index, physical activity levels or exercise habits. Methods, SOT: Please read the Neurocom manual to confirm the number of transducers (4 or 5?) and the functions of the transducers (measure horizontal and vertical forces?) Methods, SOT: Include the sensory information available for each SOT condition. Methods, SOT: State explicitly the "outcome measures" - I think they are the condition-specific equilibrium scores, composite equilibrium score (not reported in the results table but somehow mentioned in the discussion?), condition-specific strategy scores, visual ratio, somatosensory ratio and vestibular ratio. Explain these outcomes briefly. Statistical analysis: A multivariate analysis of variance may be more appropriate? Statistical analysis: The Spearman's rho test was used....and the conditions of the SOT "in the FM group". Results: Report the demographics (e.g., body mass index and physical activity level) of the subjects. Results: Analyse and report the 3 sensory ratios (visual, vestibular and somatosensory ratios) to answer the first research question. Results: State explicitly the exact outcome measures, instead of using vague terms such as "poorer postural control", "lower values", "balance responses", "the scores". Discussion, second paragraph: "Regarding the sensory analysis..." - no sensory analysis was reported in the results section? Discussion: Please explain EACH phenomenon (result). For example, WHY the FM group demonstrated lower SOT strategies scores? Discussion: Discuss the strategy score results. Discussion, limitation: Perform a post hoc power analysis to show that the current sample is small (study is under powered). I sincerely hope that the above suggestions help. I think the manuscript/data is publishable after revision. Nice to read this interesting study.

REVIEWER	Kim Dupree Jones Oregon Health & Science University
REVIEW RETURNED	24-Feb-2017

GENERAL COMMENTS	This is a well conducted and well written pilot study on women w FM and matched controls. I have no major concerns. I have 2 suggestions for the authors and 1 comment. 1) one of the study aims was to determine the relative contribution of somatsosenory, visual and vestibular to balance (SOT COMP) in people with FM. The authors state that this has not been previously reported. In fact it was reported by Jones et al. who contrary to their hypothesis that balance deficits would be driven by somatosensory loss, they instead found that all three components were significantly impaired, much like is seen in aging. 2) the authors state that a limitation may be depression/anxiety. I suggest that the authors have that data on the FIQ and could

	determine if mood were related to worse scores. comment: Lastly, the most significant contribution of this paper is the extent of functional independence testing and correlation with balance scores and disability.

REVIEWER	Julius Sim Keele University, UK
REVIEW RETURNED	18-Mar-2017

GENERAL COMMENTS	It would be helpful if a number of statistical issues were clarified and amended as appropriate: Page 6, line 40: it is stated that participants were matched on age and hand dominance. What categories of age were used for the matching? Page 8, line 14: if items were scored 1-7 and there were 18 items, surely the range of possible scores is 18 to 126, not 0 to 126? Page 9, line 40: I don't think this was a mixed model. A mixed model contains both fixed and random effects, which does not seem to be the case here. If subjects were matched, they are effectively paired: would this not in effect create two within-subjects factors: one representing the matched subjects and one representing the conditions? Page 9, line 51: was this a two-tailed alpha? If so, specify. Similarly at the top of page 10. The Bonferroni correction is stated in terms of six independent-samples t tests (though see comment above regarding the appropriateness of such between-subjects analyses). I cannot see in the results where such tests have been employed. An omnibus test seems to have been reported for the comparison of groups across the conditions on page 10 (a single p value is given as p=0.005), but it is not clear where separate condition-specific test results are. Or is this p value meant to refer to these six tests? If so, unless each p value was exactly 0.005, you need to state something like 'p<0.005 in each case'. Furthermore, if these are the comparisons in each of the conditions, their performance dependent on the significance of the omnibus test across conditions? Similar comment for the results on page 11. Page 10, line 16: 29 women with FMS were approached and 20 recruited (who were then matched with 20 healthy controls). How was this sample size determined? Was any sort of sample size calculation performed? Page 11, line 3: it is stated that a significant interaction was found. This indicates that the magnitude of the between-group differences differed across the conditions. It is then stated that values for conditions 4-6 were lower than those for conditions 1-3, and a p value of 0.007 is quoted. This appears to be a within-subjects contrast (without reference to the groups). How was this performed performed, and how, if at all, does it relate to the interaction (as

	opposed to the main effect of 'condition')? Page 11, line 54: Spearman's correlation requires a monotonic, not necessarily a linear, relationship. Unless all of the correlations were in fact linear, I would not refer to them as such. It appears that, in total, 264 correlations have been performed and tested for significance. This seems very indiscriminate and does not seem to reflect individual prior hypotheses. Furthermore it is odd that concern was shown for Type 1 error inflation in respect of the six comparisons described in conjunction with the ANOVA, but not here. If this number of correlations has to be performed, I would simply present the coefficients and not test them for significance. You can then discuss the overall pattern of the correlations in terms of their magnitude, without inviting the reader to assess individual coefficients for significance. Alternatively, could you correlate the conditions with the total FIM score, or for specific domain scores within the FIM, if they exist? Page 12: in the Discussion, as well as the Results, there is a large emphasis on the significance of the various comparisons but very little attention to their magnitude.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 - Borja Sañudo

General comment: Thank you for the opportunity to review this paper. The study is oriented to gain insight into the association between functional independence and balance responses in patients with fibromyalgia.

Response: We would like to thank to the reviewer for this positive comment

Comment 1: My main concern is the sample size. In my opinion, is too small for a cross-sectional observational study.

Response: We agree with the reviewer that 20 patients per groups could be considered a small sample size. We have included our a priori sample size calculation as follows:

"Sample Size Calculation
The sample size was calculated using the Ene 3.0 software (Autonomic University of Barcelona, Spain). The sample calculation was based on detecting significant moderate correlations (r=0.60) between the SOT conditions and FIM variables with an alpha level (α) of 0.05, and a desired power (β) of 90%. This generated a sample size of at least 19 subjects."

In addition, the topic of the sample size is also included in the limitations as follows:

"This study presents several limitations. First, although significant differences were found between both groups, these were based on a small sample size. Nevertheless, we believe that a large sample size would not alter the direction of our findings. Further, the population included was recruited from a regional hospital, which makes generalization of the results to the general population difficult. Consequently, further epidemiological studies with larger sample sizes are needed to enable a more

generalized interpretation of the results."
In addition, some points that the authors should consider which may strengthen the manuscript:

Comment 2: Abstract

Please include the meaning of SOT in the abstract.

Response: We have defined the outcomes as requested in the abstract as follows:

"The Sensory Organization Test (SOT) was used to determine postural sway and balance during six different conditions with subjects in a standing position. The Functional Independence Measure (FIM) was used to determine the level of functional independence in daily life activities"

Please include data of the correlations in the results section.

Response: We have included this data as requested:

"…Different linear correlations between balance and specific daily activities were observed within the FMS group: bathing activity and balance condition 6 ($rs=0.541$; $P<0.001$), bed transfers activity and conditions 2 ($rs=0.491$; $P<0.001$) and 3 ($rs=0.510$; $P<0.001$), positioning strategy 6 and dressing the upper ($rs=0.530$; $P<0.001$), or lower ($rs=0.562$; $P<0.001$) body, and toileting ($rs=0.521$; $P<0.001$): the greater the loss of balance, the greater the interference on some daily life activities."

Comment 3: Introduction

Overall, it is easy to follow with logic and rationale for fibromyalgia syndrome.

Response: We would like to thank to the reviewer for this positive comment

In the introduction section you justify the "poor posture" based on pain severity and fatigue and further you discuss the effects on the functional capacity of the upper extremity. This section must be refocusing on the aim of the study.

Response: This sentence has been edited as follows:

"…Finally, FM can be associated with general inactivity12 which can lead to negative effects on the functional capacity of the patient…"

However, the rationale does not seem strong enough for the association between functional independence and balance responses in this patients group. If Muto et al. (2015) and Jones et al. (2009) have previously assessed the relationship between postural control and balance self efficacy in women with fibromyalgia, what does this study add? The need for this study must be clearly discussed to the reader. Further, your aims were previously solved in different studies (e.g., Russek and Fulk, 2009), which should be cited.

Response: We have extensively rewritten the introduction section including findings from these previous studies and the new study suggested by the reviewer. We believe that the current version clearly shows the rational for this study and the differences with previous studies.

“..Some previous studies have reported the presence of balance and postural control deficits in women with FM using different procedures16-18. Muto et al17 observed that patients with FM exhibited impaired postural control, e.g., increased speed of oscillation of the center of gravity and lower balance self-efficacy as assessed with the modified clinical test of sensory interaction on balance (mCTSIB) and the balance self-efficacy (ABC scale). In this study, impaired postural control and low balance self-efficacy were associated with pain severity and muscle strength17. Jones et al16 found that FM patients showed lower scores in almost all conditions of the SOT and an increased number of falls. In this study, postural stability was associated to related disability, cognitive impairment and body mass index, but not to medication intake, pain severity or muscle strength16. In a pilot study using the SOT, Russek and Fulk18 reported that 34% of FM subjects scored below the fifth percentile for population normative data in some SOT conditions. These authors also found a negative association between the somato-sensory score of the SOT and FM-related disability18. Although these studies support the occurrence of balance problems in patients with FM using the SOT, they did not investigate the association of balance disturbances with functional independence in activities of daily living (ADLs). The identification of an association between balance problems and ADL disturbances can help clinicians for developing specific therapeutic strategies for patients with FMS. To the best of the author’s knowledge, no study has previously investigated this association in patients with FM…”

Comment 4: Methods

Please justify why the new criteria in the FM were not used.

Response: This is not really correct. In fact, we used a combination of both criteria as it has been demonstrated it is more accurate. We recognize that this was not clear in the previous version of the text. We have now clarified this issue in the methods section as follows:

“..Women with diagnosis of FM were recruited from the Department of Rheumatology at the Hospital Fundación Alcorcón (Spain). An experienced rheumatologist confirmed the FM diagnosis based on a combination of both American College of Rheumatology criteria (1990/m2010)19,20. It has been suggested that a combination of 1990 and m2010 criteria is recommended since it had the best diagnostic features21,22. Tender points were tested by digital palpation at the 18 sites according to the ACR proto¬col19. Participants were asked to indicate whether they experienced pain in response to a pressure of approximately 4 kg exerted by the examiner19. Further, the presence of fatigue, altered sleep patterns and other sensory symptoms self-perceived by the patient was recorded20. Face-to-face structured medical inter¬views were performed to determine the time of the diagnosis, socio-demographic and clinical data, current medication intake and presence of psychiatric disorders...”

[21] Segura-Jiménez V, Aparicio VA, Álvarez-Gallardo IC, Soriano-Maldonado A, Estévez-López F, Delgado-Fernández M et al. Validation of the modified 2010 American College of Rheumatology diagnostic criteria for fibromyalgia in a Spanish population. Rheumatology 2014; 53: 1803-11.

[22] Segura-Jiménez V, Soriano-Maldonado A, Álvarez-Gallardo IC, Estévez-López F, Carbonell-Baeza A, Delgado-Fernández M. Subgroups of fibromyalgia patients using the 1990 American College of Rheumatology criteria and the modified 2010 preliminary diagnostic criteria: the al-Ándalus project. Clin Exp Rheumatol 2016; 34: S26-33.

Depression is a common symptom in FM patients; if you exclude this population you may bias your results. Please report this issue.

Response: We agree with the reviewer that depression is highly comorbid with FM. We decided to exclude these patients for avoiding any influence on the outcomes. We have included in the limitation section this issue as follows:

"…Third, we excluded women with FM and comorbid depressive symptoms, so extrapolation of our results to this subgroup of patients with FM should be considered with caution. Although it seems that depression or anxiety may affect balance37; we do not known the effect of depression in the outcomes included in our study, particularly those related to the FIM…"

"Each condition was performed 3 consecutive times." Please indicate how the results were computed and included for further analysis.

Response: We have clarified that the mean of the 3 trials was used in the analysis:

"…Each condition was performed 3 consecutive times and the mean was considered in the analysis…"

Statistical analysis: Did you check for normality?

Response: Yes, of course. We have included this info in the statistical analysis section as follows:

"…The Kolmogorov-Smirnov test was used to analyze the normal distribution of the variables (P>0.05). Quantitative data without a normal distribution (clinical data and FIM scores) were analyzed with non-parametric tests and those data with a normal distribution (SOT conditions) were analyzed with parametric tests..."

Was the sample size for a Spearman correlation determined using power analysis? Generally a minimum of 29 participants must to be recruited to determine a reasonable high correlation of two variables (to be able to detect effect size to detect correlation coefficient = 0.5 with alpha and power are set at 0.05 and 80% respectively).

Response: We have included our a priori sample size calculation as follows:

"Sample Size Calculation

The sample size was calculated using the Ene 3.0 software (Autonomic University of Barcelona, Spain). The sample calculation was based on detecting significant moderate correlations (r=0.60) between the SOT conditions and FIM variables with an alpha level (α) of 0.05, and a desired power (β) of 90%. This generated a sample size of at least 19 subjects."

Comment 5: Discussion

The discussion should be focused towards your true findings rather than what others have found previously. While that is important to give context to the findings, the reviewer wasn't sure of what your findings mean and in which context. Further, witch such low correlations I believe that your statements must be modulated.

Response: We have modulated the tone of our findings as the reviewer has suggested and we have based our discussion in the main findings of the study, but obviously we should compare with

previous studies.

On page 12 you report that “lower scores were observed in conditions 4 to 6, which are the more challenging conditions, as these correspond with the mobile platform conditions.” But this is also the case in control participants, thus, I´m not sure if your discussion on this topic is suitable.

Response: We believe that the discussion of the vestibular system is important. In fact, conditions 4-6 are more related to vestibular system than conditions 1-3. Since the vestibular ratio score was the only significantly impaired in our sample of women with FM. We have rewritten this section as follows:

“…Women with FM exhibited lower scores in all SOT conditions compared to healthy women suggesting poor balance. Our results agree with those previously reported by Jones et al16 and Russek and Fulk18 who also reported significantly lower scores in all SOT conditions in individuals with FM. It is interesting to note that the scores observed in our study were similar to those reported in these previous studies16,18. Current and previous evidence would suggest that subjects with FM exhibit poor general balance as compared to healthy women. Nevertheless, although all SOT conditions showed lower scores in FM, the vestibular ratio, but not the somato-sensory or visual ratio, was the most significantly impaired in our sample of women with FM. This may be related to the fact that the scores in the last SOT conditions (4 to 6) were significantly much lower in the FM group than in the healthy group. As previously suggested, lower scores in conditions 4 to 6 compared to conditions 1 to 3 suggest a degree of somato-sensory dependence15. This hypothesis is in line with the study by De Brujin et al33 who found that balance in patients with FM was most optimal on firm and regular surfaces. In fact, Russek and Fulk18 and our study did not find significant differences within the somato-sensory system ratio between patients with FM and healthy people, suggesting that it is the vestibular, and probably the visual, system16,18 the most affected in this population.”

To justify the fact that there is no consensus in the management of these patients you use very old references. In the last couple of years there are many reviews and meta-analysis dealing with this issue. I suggest rewriting this paragraph. The authors should mention the implications for patients’ management if any, but on the view of their results.

Response: We have edited this paragraph and also the discussion section. We have now deleted this sentence about the consensus on management of these patients since this is not related to our study and we have included recommendation for management of these patients based on our results as follows:

“…The most relevant result of our study was the positive association between balance scores and functional independence during ADL since the greater the loss of postural balance, the greater the interference with those ADL activities requiring proper postural control and balance, e.g., bathing and dressing. These findings are valuable for planning proper treatment interventions, since deficits or loss of independence in these ADLs has a negative impact on quality of life of the patients. Our results agree with the study by Amris et al36 who investigated 257 women with widespread chronic pain symptoms and observed that patients with FM have substantial problems affecting their daily life and are liable to need community support. Therefore, current findings can help for planning multidisciplinary interventions for individuals with FM. For instance, balance strategies and postural control can be treated with physical therapy whereas therapeutic strategies for improvement of ADL efficacy should be applied by occupational therapists. Further, cognitive behaviors or fear to movement can be benefit from psychological approaches”

Again on page 13 authors stated that this is the first study assessing the relationship between balance and functional independence in ADLs in this population group and there are a few previous studies that have not been discussed (e.g., Goes, 2012 or some preliminary results from The al-Ándalus project)

Response: We would like to comment to the reviewer that this is the first study assessing the relationship between balance and ADL. The study by Goes et al published in Clinical Biomechanics in 2012 did not assess ADL. This study assessed balance with muscle function and falls, but not ADL. We believe that we have included the most relevant papers on the topic of falls considering that our study did not investigate falls in FM. Similarly, we also know the studies published in the Al-Andalus project, but most of them are related to physical activity and fitness and not balance. We have carefully checked all published studies and there are no references to balance or ADL using the outcomes from our study. We have included one reference from the Al-Andalus project in relation to the use of combined diagnostic criteria for FM, one of the comments from this reviewer (see previous comment in methods)

The physiological basis of the deficits observed in these patients must to be discussed in detail.

Response: We have included some comments on physiological basis for the deficits. We have now expanded too much this new section since we believe this would be highly speculative.

"…To determine the mechanisms related to poor balance in patients with FM is beyond the scope of the current study, but some hypotheses have been proposed. Since FM is characterized by abnormal nociceptive processing, it is possible that multiple processing disturbances may lead to poor balance. Additionally, other processing abnormalities of the central nervous system, e.g., cognitive dysfunction, could also contribute to postural instability. In fact, Bayazit et al34 suggested that women with FM have neural brainstem disintegration which could lead to abnormal perception of audio-vestibular inputs and to an abnormal auditory brainstem response. Current and previous finding demonstrating that the vestibular system was the most affected in individuals with FM would support this hypothesis. Nevertheless, because we did not specifically assess the function of the vestibular system in our sample of women with FM, the current results do not determine whether the low scores on the vestibular component of the SOT were due to peripheral or central deficits…"

It would be interesting to look at the differences regarding symptom severity.

Response: As it is stated in the results section, no associations between balance scores and clinical outcomes including symptoms severity were found.

In the limitations authors suggest that the selection of your sample "makes generalization of the results to the general population difficult." However, they conclude on the general FM population. This must be modulated.

Response: We have modulated the tone of our findings as the reviewer has suggested.

How can you suggest multidisciplinary treatments based on your results?

Response: We have included the following comment on this topic:

"…The most relevant result of our study was the positive association between balance scores and functional independence during ADL since the greater the loss of postural balance, the greater the interference with those ADL activities requiring proper postural control and balance, e.g., bathing and dressing. These findings are valuable for planning proper treatment interventions, since deficits or loss of independence in these ADLs has a negative impact on quality of life of the patients. Our results agree with the study by Amris et al36 who investigated 257 women with widespread chronic pain symptoms and observed that patients with FM have substantial problems affecting their daily life and are liable to need community support. Therefore, current findings can help for planning multidisciplinary interventions for individuals with FM. For instance, balance strategies and postural control can be treated with physical therapy whereas therapeutic strategies for improvement of ADL efficacy should be applied by occupational therapists. Further, cognitive behaviors or fear to movement can be benefit from psychological approaches"

Reviewer 2 - Shirley S.M. Fong

Comment 1: Abstract, Conclusion: ..."resulting in" negative impact... Better use the term "associated with" negative impact.

Response: Edited

Comment 2: Introduction:

Aims of study: to investigate which sensory opponent is the most affected (vestibular, visual or somato-sensory)...However, the sensory organization test vestibular, visual and somato-sensory ratios were not reported in the manuscript. Must report the 3 sensory ratios

Response: We have now included the vestibular, visual and somato-sensory score ratios as requested in the methods and results sections as requested:

"…In addition, combination of the results obtained in the different conditions provides a ratio score of each sensory system (somato-sensory, vestibular, or visual). The somato-sensory ratio (condition 2/condition 1) determines how successfully a person uses input from the somato-sensory system for balance; the visual ratio (condition 4/condition 1) determines how successfully a person uses visual system for balance; and the vestibular ratio (condition 5/condition 1) determines how successfully a person uses input from the vestibular system for balance…

..Further, unpaired Student t-tests were also conducted to determine between-groups differences for the ratio score of each sensory system (somato-sensory, vestibular, or visual)…

… We found significant (t=2.901; P=0.006) lower vestibular ratio score in women with FM (mean: 0.55±0.2) as compared to healthy women (mean: 0.72±0.15). No significant differences in somato-sensory (t=0.011; P=0.989) and visual (t=1.900; P=0.065) ratios between women with FM (somato-sensory: 0.95±0.03; visual: 0.82±0.15) and healthy women (somato-sensory: 0.96±0.03; visual: 0.90±0.1) were observed…."

Aims of study: With the current results, I think another aim of the study was to compare the sensory organization of balance control and balance strategies between women with FM and controls (i.e., the two-way repeated measures group effect). The time and interaction effects were also the aims of the study?

Response: The reviewer is quite right. We have edited the objective section as follows:

“Therefore, the aims of the current study were: 1) to compare sensory organization of balance control and balance strategies between women with FM and healthy controls; 2) to investigate which sensory component (vestibular, visual or somato-sensory) is the most affected in FM women; and, 3), to determine the potential association between the functional independence measure (FIM) and balance responses in women with FM.”

Comment 3: Methods:

Inclusion criteria: instead of stating "NO pain" and "NO history" (move these to the exclusion criteria), should include age range and other demographics such as body mass index, physical activity levels or exercise habits.

Response: Since these criteria are related to the healthy control group, we believe that the absence of pain is an inclusion criterion, not an exclusion criterion. We have edited this paragraph as follows:

“…Participants were considered as healthy controls if they reported: no spontaneous pain symptoms at the moment of the study, no history of chronic pain (lasting more than 3 months), no pain experienced during the previous year prior to the study, no pain-related diagnoses and participants who were not tak¬ing antidepressant or analgesic medication...”

Demographic data are included in the results section. We have now included body mass index. We did not recorded physical activity levels.

“…There were no significant dif¬ferences in age (P=0.909) or body mass index (control: 23.8 ±1.3; FM: 24.2±1.5, P=0.508) between both groups…”

Finally, we have also included the BMI as a covariate for the main analysis as follows:

“…A 2x6 analysis of variance (ANCOVA) with group (FM or controls) as a between-subjects factor and with condition of the SOT (from 1 to 6) as a within-subjects factor was and body mass index as covariate used to analyze differences in the assessments of balance responses and strategies used for maintaining the upright position in the SOT...”

And the results section:

… The ANCOVA revealed significant differences between groups ($F=21.634$; $P<0.001$) and conditions ($F=45.164$; $P<0.001$) for the balance responses on the SOT: women with FM displayed significantly ($P=0.005$) lower values in all SOT conditions compared to healthy women (table 1). A significant Group * Condition interaction was also observed ($F=3.404$; $P=0.006$): scores of conditions 4-6 were significantly lower ($P=0.007$) than those for conditions 1-3, particularly within the FM group. No effect of the body mass index was observed ($F=1.144$; $P=0.338$)…”

SOT: Please read the Neurocom manual to confirm the number of transducers (4 or 5?) and the functions of the transducers (measure horizontal and vertical forces?)

Response: We have edited this sentence as follows:

“The device consists of a platform connected to symmetrically placed transducers measuring the vertical and horizontal shear forces exercised through the anterior-posterior axis in the plane parallel to the floor”

SOT: Include the sensory information available for each SOT condition.

SOT: State explicitly the "outcome measures" - I think they are the condition-specific equilibrium scores, composite equilibrium score (not reported in the results table but somehow mentioned in the discussion?), condition-specific strategy scores, visual ratio, somato-sensory ratio and vestibular ratio. Explain these outcomes briefly.

Response: This section has been completely rewritten as follows:

“To conduct the SOT, an individual’s postural sway, and thereby balance is measured under 6 different conditions during standing. During these tests, the base of support and the visual surround screen can move according to the patient’s balancing responses and the strategy used for maintaining an upright position. For instance, no altered stimuli are given in condition 1; whereas visual information is removed in condition 2, by asking the participant to close the eyes. In condition 3, the visual surround is moving with the subject’s anterior-posterior body sway, whereas in condition 4, the platform rotates with the subject’s anterior-posterior body sway. In condition 5, subjects close their eyes and the platform moves with the subject anterior-posterior body sway. Finally, in condition 6, the visual screen and the platform are moved with the subject’s anterior-posterior body sway. Briefly, the 6 conditions can be resumed as follows: 1) eyes open, fixed surround and support platform; 2) eyes closed, fixed surround and support platform; 3) eyes open, moving surround (moving proportional to the angle of anterior-posterior body sway) and fixed support platform; 4) eyes open, fixed surround and moving support platform (moving proportional to the angle of anterior-posterior body sway); 5) eyes closed, fixed surround and moving support platform; and 6) eyes open, moving surround and support platform. Tests were always performed following these steps in order. Each condition was performed 3 consecutive times and the mean was considered in the analysis. In total, the duration of the tests lasted approximately 12 minutes for each patient; therefore, it can be considered a non-fatiguing assessment. This procedure has shown good test-retest reliability in healthy people31.

Participants were encouraged to maintain their stability and center of gravity, despite the movement of the visual surround or the base of support. The participant’s center of gravity was displayed on the upper half of the screen. The feet were correctly positioned facing the visual surround during the entire test. If the participant fell, took a step or touched the visual surround, the test was interrupted and the fall was registered. Data assessments were performed automatically and compared with theoretical normative electronic data. The score of each condition consist of a percentage that compares the subject anterior-posterior center of pressure sway with the theoretical limits of stability. The score is registered on a bar chart ranging from 0% to 100% where 0% represents the least stable (fall) and 100% indicates perfect stability28.

In addition, combination of the results obtained in the different conditions provides a ratio score of each sensory system (somato-sensory, vestibular, or visual). The somato-sensory ratio (condition 2/condition 1) determines how successfully a person uses input from the somato-sensory system for balance; the visual ratio (condition 4/condition 1) determines how successfully a person uses visual system for balance; and the vestibular ratio (condition 5/condition 1) determines how successfully a person uses input from the vestibular system for balance.

Finally, a strategy score for each SOT condition is also calculated with scores near 100 indicating use of an ankle strategy and scores near 0 indicating a hip strategy."

Statistical analysis: A multivariate analysis of variance may be more appropriate?

Response: We have consulted with our experienced statistician and he recommends us the ANOVA due to the nature of the study and the sample size.

Statistical analysis: The Spearman's rho test was used....and the conditions of the SOT "in the FM group".

Response: Edited

Comment 4: Results:

Report the demographics (e.g., body mass index and physical activity level) of the subjects.

Response: Demographic data are included in the results section. We have now included body mass index. We did not recorded physical activity levels.

"…The final sample consisted of 20 women with FM aged 35-55 years old (mean: 48±6 years) who satisfied all the eligibility criteria and agreed to participate. In addition, 20 matched healthy women; aged 35-56 years old (mean: 47±6 years) were also included. There were no significant dif¬ferences in age (P=0.909) or body mass index (control: 23.8 ±1.3; FM: 24.2±1.5, P=0.508) between both groups. All participants were right-handed. Seventeen (85%) women with FM (85%) were regularly taking non-steroidal anti-inflammatory medications. The FIQ revealed a moderate disability with a mean score of 57.9 (95%CI 53.1-62.6)…"

Analyze and report the 3 sensory ratios (visual, vestibular and somato-sensory ratios) to answer the first research question.

Response: We have now included the vestibular, visual and somato-sensory score ratios as requested in the methods and results sections as requested:

"…In addition, combination of the results obtained in the different conditions provides a ratio score of each sensory system (somato-sensory, vestibular, or visual). The somato-sensory ratio (condition 2/condition 1) determines how successfully a person uses input from the somato-sensory system for balance; the visual ratio (condition 4/condition 1) determines how successfully a person uses visual system for balance; and the vestibular ratio (condition 5/condition 1) determines how successfully a person uses input from the vestibular system for balance…

..Further, unpaired Student t-tests were also conducted to determine between-groups differences for the ratio score of each sensory system (somato-sensory, vestibular, or visual)…

… We found significant (t=2.901; P=0.006) lower vestibular ratio score in women with FM (mean: 0.55±0.2) as compared to healthy women (mean: 0.72±0.15). No significant differences in somato-sensory (t=0.011; P=0.989) and visual (t=1.900; P=0.065) ratios between women with FM (somato-sensory: 0.95±0.03; visual: 0.82±0.15) and healthy women (somato-sensory: 0.96±0.03; visual:

0.90±0.1) were observed…."

State explicitly the exact outcome measures, instead of using vague terms such as "poorer postural control", "lower values", "balance responses", "the scores".

Response: We have specified the proper terms in the results section as requested.

Comment 5: Discussion

Second paragraph: "Regarding the sensory analysis..." - no sensory analysis was reported in the results section?

Response: The discussion section has been extensively rewritten.

Please explain EACH phenomenon (result). For example, WHY the FM group demonstrated lower SOT strategies scores?

Response: We have included some comments on physiological basis for the deficits. We have now expanded too much this new section since we believe this would be highly speculative.

"…To determine the mechanisms related to poor balance in patients with FM is beyond the scope of the current study, but some hypotheses have been proposed. Since FM is characterized by abnormal nociceptive processing, it is possible that multiple processing disturbances may lead to poor balance. Additionally, other processing abnormalities of the central nervous system, e.g., cognitive dysfunction, could also contribute to postural instability. In fact, Bayazit et al34 suggested that women with FM have neural brainstem disintegration which could lead to abnormal perception of audio-vestibular inputs and to an abnormal auditory brainstem response. Current and previous finding demonstrating that the vestibular system was the most affected in individuals with FM would support this hypothesis. Nevertheless, because we did not specifically assess the function of the vestibular system in our sample of women with FM, the current results do not determine whether the low scores on the vestibular component of the SOT were due to peripheral or central deficits…"

Discuss the strategy score results.

Response: We have included a new paragraph with this topic as follows:

"…Additionally to lower balance scores, we also observed that our sample of women with FM also used different strategy than healthy women for maintaining their balance. The SOT strategy scores indicate that woman with FM use a hip strategy to maintain their balance whereas healthy women use a more ankle strategy. Some possible reasons for this change in balance strategy can be the presence of muscle trigger points in the gastrocnemius and tibialis anterior muscles16 or the greater muscle fatigue in the tibialis anterior muscle35. Future studies should investigate neurophysiological mechanisms related to changes in balance strategy in subjects with FM…"

Limitation: Perform a post hoc power analysis to show that the current sample is small (study is under powered).

Response: The topic of the sample size has been previously answered to the Reviewer 1. We have provided our a priori sample size calculation and also a comment in the limitation section as

requested.

I sincerely hope that the above suggestions help. I think the manuscript/data is publishable after revision. Nice to read this interesting study

Response: We would like to thank to the reviewer for this positive comment

Reviewer 3 - Kim Dupree Jones

This is a well conducted and well written pilot study on women w FM and matched controls. I have no major concerns.

Response: We would like to thank to the reviewer for this positive comment

I have 2 suggestions for the authors and 1 comment.

Comment 1: one of the study aims was to determine the relative contribution of somato-sensory, visual and vestibular to balance (SOT COMP) in people with FM. The authors state that this has not been previously reported. In fact it was reported by Jones et al. who contrary to their hypothesis that balance deficits would be driven by somato-sensory loss, they instead found that all three components were significantly impaired, much like is seen in aging.

Response: We have extensively rewritten the introduction section including findings from these previous studies and the new study suggested by the reviewer. We believe that the current version clearly shows the rational for this study and the differences with previous studies.

"..Some previous studies have reported the presence of balance and postural control deficits in women with FM using different procedures16-18. Muto et al17 observed that patients with FM exhibited impaired postural control, e.g., increased speed of oscillation of the center of gravity and lower balance self-efficacy as assessed with the modified clinical test of sensory interaction on balance (mCTSIB) and the balance self-efficacy (ABC scale). In this study, impaired postural control and low balance self-efficacy were associated with pain severity and muscle strength17. Jones et al16 found that FM patients showed lower scores in almost all conditions of the SOT and an increased number of falls. In this study, postural stability was associated to related disability, cognitive impairment and body mass index, but not to medication intake, pain severity or muscle strength16. In a pilot study using the SOT, Russek and Fulk18 reported that 34% of FM subjects scored below the fifth percentile for population normative data in some SOT conditions. These authors also found a negative association between the somato-sensory score of the SOT and FM-related disability18. Although these studies support the occurrence of balance problems in patients with FM using the SOT, they did not investigate the association of balance disturbances with functional independence in activities of daily living (ADLs). The identification of an association between balance problems and ADL disturbances can help clinicians for developing specific therapeutic strategies for patients with FMS. To the best of the author's knowledge, no study has previously investigated this association in patients with FM…"

Comment 2: the authors state that a limitation may be depression/anxiety. I suggest that the authors have that data on the FIQ and could determine if mood were related to worse scores.

Response: As we have previously commented to previous reviewers, we excluded patients with mood disorders. This is included in the limitation section.

Comment 3: Lastly, the most significant contribution of this paper is the extent of functional independence testing and correlation with balance scores and disability.

Response: We agree with the reviewer that this is the most relevant contribution of our paper. This has been clarified in the discussion. In fact, the sample size calculation was a priori based on these associations as it has been clarified in the current version of the paper.

Reviewer 4 - Julius Sim

It would be helpful if a number of statistical issues were clarified and amended as appropriate:

Response: We have answered to all comments provided by the reviewer, thanks

Comment 1: Page 6, line 40: it is stated that participants were matched on age and hand dominance. What categories of age were used for the matching?

Response: There were no age categories; we matched patient by patient and age by age. This is the common age matching procedure.

Comment 2: Page 8, line 14: if items were scored 1-7 and there were 18 items, surely the range of possible scores is 18 to 126, not 0 to 126?

Response: The reviewer is quite right, thanks for this typo. This has been edited.

"…The possible score ranges from 18 to 126 points…"

Comment 3: Page 9, line 40: I don't think this was a mixed model. A mixed model contains both fixed and random effects, which does not seem to be the case here.
If subjects were matched, they are effectively paired: would this not in effect create two within-subjects factors: one representing the matched subjects and one representing the conditions?

Response: We have deleted the words "mixed-model" since it may be confusing. We have consulted with our statistician and he confirmed us that subjects, although they were matched, it is a between-subjects conditions since there are different subjects on each group. In fact, due to some comments from previous reviewers, we have conducted an ANCOVA since we have now included BMI in the analysis as covariate.

Comment 4: Page 9, line 51: was this a two-tailed alpha? If so, specify. Similarly at the top of page 10

Response: This sentence has been edited according to the following comment.

Comment 5: The Bonferroni correction is stated in terms of six independent-samples t tests (though see comment above regarding the appropriateness of such between-subjects analyses). I cannot see in the results where such tests have been employed. An omnibus test seems to have been reported for the comparison of groups across the conditions on page 10 (a single p value is given as p=0.005), but it is not clear where separate condition-specific test results are. Or is this p value meant to refer to these six tests? If so, unless each p value was exactly 0.005, you need to state something like 'p<0.005 in each case'. Furthermore, if these are the comparisons in each of the conditions, their performance dependent on the significance of the omnibus test across conditions? Similar comment for the results on page 11

Response: The Bonferroni correction used in the current study is not a post hoc analysis as we believe the reviewer is referring to. We were very restrictive with our statistical significance and we use the Holm-Bonferroni adjustment. We recognize that the use of correction can be confused. We have rewritten this section of the statistical analysis as follows:

"The statistical analysis was conducted at a 95% confidence level; but, we corrected for multiple

comparisons using the Holm-Bonferroni adjustment32 assuming a significant alpha level of 0.008 (6 in¬dependent-samples t-tests by condition).”

[31] Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979; 6: 65-70

Comment 6: Page 10, line 16: 29 women with FMS were approached and 20 recruited (who were then matched with 20 healthy controls). How was this sample size determined? Was any sort of sample size calculation performed?

Response: We have included our a priori sample size calculation as follows:

“Sample Size Calculation
The sample size was calculated using the Ene 3.0 software (Autonomic University of Barcelona, Spain). The sample calculation was based on detecting significant moderate correlations (r=0.60) between the SOT conditions and FIM variables with an alpha level (α) of 0.05, and a desired power (β) of 90%. This generated a sample size of at least 19 subjects.”

Comment 7: Page 11, line 3: it is stated that a significant interaction was found. This indicates that the magnitude of the between-group differences differed across the conditions. It is then stated that values for conditions 4-6 were lower than those for conditions 1-3, and a p value of 0.007 is quoted. This appears to be a within-subjects contrast (without reference to the groups). How was this performed, and how, if at all, does it relate to the interaction (as opposed to the main effect of ‘condition’)?

Response: We agree with the reviewer that this paragraph is slightly confused as it was written. The Group*Condition interaction only says that differences between conditions were more pronounced within the FM group. We have edited this section as follows:

“...The ANCOVA revealed significant differences between groups (F=21.634; P<0.001) and conditions (F=45.164; P<0.001) for the balance responses on the SOT: women with FM displayed significantly (P=0.005) lower values in all SOT conditions than healthy women and scores of conditions 4-6 were significantly lower (P=0.007) than those for conditions 1-3 (table 1). A significant Group * Condition interaction was also observed (F=3.404; P=0.006): differences between conditions 4-6 and conditions 1-3 were more pronounced within the FM group”

Comment 8: Page 11, line 54: Spearman’s correlation requires a monotonic, not necessarily a linear, relationship. Unless all of the correlations were in fact linear, I would not refer to them as such.

Response: We have removed the word linear as suggested.

Comment 9: It appears that, in total, 264 correlations have been performed and tested for significance. This seems very indiscriminate and does not seem to reflect individual prior hypotheses. Furthermore it is odd that concern was shown for Type 1 error inflation in respect of the six comparisons described in conjunction with the ANOVA, but not here. If this number of correlations has to be performed, I would simply present the coefficients and not test them for significance. You can then discuss the overall pattern of the correlations in terms of their magnitude, without inviting the reader to assess individual coefficients for significance. Alternatively, could you correlate the conditions with the total FIM score, or for specific domain scores within the FIM, if they exist?

Response: We agree with the reviewer that we conducted several correlations. As the reviewer can pointed out, we have only considered really significant in the text those with P values lower than 0.008 as determined by the Bonferroni adjustment. We have clarified in the text of the statistical analysis that the adjustment on significance was also applied to the correlational analysis. We believe that we should maintain their P values since not all were significant and discussed.

Comment 10: Page 12: in the Discussion, as well as the Results, there is a large emphasis on the significance of the various comparisons but very little attention to their magnitude.

Response: The discussion has been extensively rewritten. It is very difficult to discuss the magnitude of the differences since there are no cut off scores on the outcomes. We believe that the current version of the discussion is much focused on the main results of the study.

We hope that the current version satisfies all comments from the reviewers and can be accepted for publication in BMJ Open

VERSION 2 – REVIEW

REVIEWER	Shirley S.M. Fong University of Hong Kong
REVIEW RETURNED	07-Apr-2017

GENERAL COMMENTS	It's fine now. Thanks.

REVIEWER	Julius Sim Keele University, UK
REVIEW RETURNED	14-May-2017

GENERAL COMMENTS	Thank you for clarifying that participants were matched by specific age (presumably by specific year), rather than in age bands. It is true that the participants were in two groups, but the fact of their being matched would normally suggest that they should be analysed using a within- rather than a between-subjects analysis. Can you explain more specifically why you did not do this? In response to the question about Bonferroni adjustments, you state that the 'statistical analysis was conducted at a 95% confidence level.' Although a 95% confidence level and a 5% significance level are equivalent, as you are discussing hypothesis tests, not interval estimates, I would suggest referring to significance, not confidence (specifying two-tailed significance). Your stated sample size calculation is presented in terms of 90% power. I think you will find that the stated figure of 19 subjects represents 80% power. 90% power would require 25 subjects. With regard to the large number of correlation coefficients (264) that are tested for significance, you state that the adjustment for significance was applied to these tests as well as to the tests for differences in means. However, the adjustment is stated in relation to 6 comparisons of means ("6 independent-samples t-tests by condition"). How does this adjustment relate to the 264 correlations?

	You also state that the p values (and hence the tests) should be retained as “not all were significant and discussed”. If a decision is made to adjust alpha, this must be with regard to the number of tests that were performed, not the number that were significant (and/or discussed). It is unclear how Type 1 error was controlled in relation to the correlations. It is still unclear how a difference was determined between conditions 1 to 3 and conditions 4 to 6 (page 12 and Table 1). Do you mean that each of conditions 4 to 6 differed significantly from each of conditions 1 to 3 (if so, how was a single p value derived from these 9 contrasts?), or does this represent a significant difference between the average of conditions 4 to 6 and the average of conditions 1 to 3? Returning to the p value, in the text this seems to be given as “p = 0.007”, whereas “p < 0.001” is quoted in Table 1. If these p values relate to the same analysis, they seem inconsistent. On page 13 you state that “no… effect of the body mass index (F=0.608; P = 0.770) was observed”. If BMI was a covariate you are presumably using it purely as a control variable, rather than a factor (“effect”) on which a hypothesis is to be tested. Its p value is therefore not of interest and need not be quoted.

VERSION 2 – AUTHOR RESPONSE

Reviewer 4 - Julius Sim

Comment 1: Thank you for clarifying that participants were matched by specific age (presumably by specific year), rather than in age bands. It is true that the participants were in two groups, but the fact of their being matched would normally suggest that they should be analyzed using a within- rather than a between-subjects analysis. Can you explain more specifically why you did not do this?

Response: We have consulted to our experienced statistician and he confirms that the proper analysis involves between-subjects since both groups, although matched, are completely different mainly based on the presence or absence of a disease, in this case, FMS. The use of within-subject analysis is not proper when there is this main difference between two groups. In fact, similar papers published use the same approach.

Comment 2: In response to the question about Bonferroni adjustments, you state that the ‘statistical analysis was conducted at a 95% confidence level.’ Although a 95% confidence level and a 5% significance level are equivalent, as you are discussing hypothesis tests, not interval estimates, I would suggest referring to significance, not confidence (specifying two-tailed significance).

Response: We have used the word significance instead of confidence as requested.

Comment 3: Your stated sample size calculation is presented in terms of 90% power. I think you will find that the stated figure of 19 subjects represents 80% power. 90% power would require 25 subjects.

Response: The reviewer is quite right. This is a typo when we passed the data from the sample size calculation program, we apologize.

Comment 4: With regard to the large number of correlation coefficients (264) that are tested for significance, you state that the adjustment for significance was applied to these tests as well as to the

tests for differences in means. However, the adjustment is stated in relation to 6 comparisons of means (“6 independent-samples t-tests by condition”). How does this adjustment relate to the 264 correlations? You also state that the p values (and hence the tests) should be retained as “not all were significant and discussed”. If a decision is made to adjust alpha, this must be with regard to the number of tests that were performed, not the number that were significant (and/or discussed). It is unclear how Type 1 error was controlled in relation to the correlations.

Response: We can observe a misunderstanding. As it is stated in the statistical analysis section, the P value for the correlation was not adjusted, since we the adjustment cannot be conducted with 264 comparisons. The P value for correlation was 0.05. We mainly used the adjusted for the 6 SOT conditions. We have clarified this on page 11as follows

“The statistical analysis was generally conducted at a 95% significance level; but, we corrected for multiple comparisons using the Holm-Bonferroni adjustment32 assuming a significant alpha level of 0.008 (6 in¬dependent-samples t-tests by SOT condition).

We have also included the topic of Type I error in the correlation coefficient analysis in the limitation section on page 16 as follows:

“Second, we analyzed around 264 correlations in our study. It is possible that a Type I error would be present. A greater sample size would help to elucidate if the significant association observed in the current study are further significant or not.”

Comment 5: It is still unclear how a difference was determined between conditions 1 to 3 and conditions 4 to 6 (page 12 and Table 1). Do you mean that each of conditions 4 to 6 differed significantly from each of conditions 1 to 3 (if so, how was a single p value derived from these 9 contrasts?), or does this represent a significant difference between the average of conditions 4 to 6 and the average of conditions 1 to 3? Returning to the p value, in the text this seems to be given as “p = 0.007”, whereas “p < 0.001” is quoted in Table 1. If these p values relate to the same analysis, they seem inconsistent.

Response: We are referring that each condition 4 to 6 was significantly different from each 1 to 3. The reviewer is quite right about the misspelling in the footnote of the table. P values were changed. P<0.001 is from the main interaction and P<0.01 is for the condition differences. We have fixed this. In relation to P values between the condition comparison, all ranged between P=0.003 and 0.009. We have therefore, include all, P<0.01 as it is not in the footnote of the table.

Comment 5: On page 13 you state that “no… effect of the body mass index (F=0.608; P = 0.770) was observed”. If BMI was a covariate you are presumably using it purely as a control variable, rather than a factor (“effect”) on which a hypothesis is to be tested. Its p value is therefore not of interest and need not be quoted.

Response: We have removed the P value as requested.

We hope that the current version satisfies all comments from the reviewers and can be accepted for publication in BMJ Open

Sincerely yours - The authors