PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Can physical assessment techniques aid diagnosis in people with chronic fatigue syndrome/myalgic encephalomyelitis? A diagnostic accuracy study.
AUTHORS	Hives, Lucy; Bradley, Alice; Richards, Jim; Sutton, Chris; Selfe, James; Basu, Bhaskar; McGuire, Kerry; Sumner, Gail; Gaber, Tarek; Mukherjee, Annice; Perrin, Raymond

VERSION 1 – REVIEW

REVIEWER	B. Puri Imperial College London, UK
REVIEW RETURNED	07-May-2017

GENERAL COMMENTS	The results of this relatively large and impressive study are well described. The authors correctly mention the fact that the two groups were not matched for sex; this does not affect the relevant values of Cohen's kappa. They rightly also draw attention to the need for further studies which examine more severely affected patients.

REVIEWER	Dr Jo Daniels University of Bath
REVIEW RETURNED	01-Jun-2017

GENERAL COMMENTS	This is a very interesting paper that is potentially of benefit to patients and clinicians, with the prospect of a more expedient and reliable method of diagnosis. Clear introduction, well written. Perhaps underplaying the current utility of the Fukuda diagnostic criteria that is used across clinical services - an accepted method for diagnosis. Context is important. I appreciate that a technique such as this would be very useful in clinical settings, however I am unconvinced by this paper. It is lacking in detail and without a comparison group and clearer explanations of methods and rationale for particular methods I do not feel it is of sufficiently publishable standard for the BMJ. I would suggest that the authors consult a statistician re potential ROC analysis - I am not adequately qualified to advise in this area, however it is worth recommending the consideration of further more sophisticated tests. For me the originality is there, but the scientific rigour is not, this may be addressed in a substantial revision to aid transparency and rigour. I thank the authors for their interesting contribution.

	Specific comments: 1. Reference line 35, author pg 2 - Health and Care excellence? Not in refs. Do you mean to refer to NICE here? 2. I understand there is a protocol to refer to, but there is insufficient information here to make sense of the method of the study. Please say more. 3. Please insert ethics numbers etc. in a more appropriate place, ethics section perhaps> 4. When you say 'circulated to the wider community' what do you mean? Please offer more detail in the procedure. 5. How did you recruit non-blood relatives exactly? The level of detail makes this nonreplicable, despite a protocol. Please make the protocol available as a supplement. 6. Inclusion/exclusion criteria please? I see you have mentioned this later, however a more traditional methods section with inclusion and exclusion would be more suitable. 7. stating 'briefly' is feels inappropriate here; as scientific paper details are required for replicability. Expand: what are you not stating? 8. What is the rationale for the 4+ inclusion and exclusion criteria? Is this evidence based? 9. Please describe/introduce the 'Perrin' technique before referring to it. 10. The authors report that there is no conversation between the participant and the examiner. Can you perhaps consider the vast literature relating to patient posture and facial expressions etc that may have unblind them? 11. I'm unconvinced of the utility of this without a comparison group, i.e how would fibromyalgia patients fair here? Or patients with HIV? 12. What level of training did the clinicians have? AHP and others. 13. The statistics appear to be fairly rudimentary here - would a ROC analysis not be appropriate? This would be an appropriately sophisticated analysis for reporting sensitivity and specificity of clinical test. 14. Please correct the reference style so that it is in line with the journal guidance, e.g. how many authors listed in text and style. 15. Please be more transparent on the consent process, this is unclear.

REVIEWER	Jiawen Zhu, statistical scientist Roche Innovation Center, USA
REVIEW RETURNED	03-Jul-2017

GENERAL COMMENTS	Introduction The authors present a study to examine the agreement of a physical assessment technique for CFS/ME. The goal of the study is to assess the 5 physical signs and comparing with the standard clinical neurological and rheumatologically examination. The authors conducted a diagnostic accuracy study with CFS/ME patients and a control group. Through sensitivity and specificity analyses of diagnosis results by three practitioners, examination of the diagnosis bias with McNemar's test and agreement between both diagnosis result and individual physical signs from experienced and newly trained AHP, the authors concluded using the physical signs appears to improve the accuracy of CFS/ME diagnosis results and only two of the physical signs may be needed based on the agreement assessment. Major comments 1, by using one practitioner in each category, the study didn't consider the within category variation by practitioners, i.e. experienced AHP, newly trained AHP and physician. This is a major limitation to interpret the research result and make the conclusion to the general level, i.e. using the physical signs improves the accuracy of identifying patient with CFS/ME. A study design including diagnosis result from several experienced AHP, newly trained AHP and physician will help to answer the research question. Minor comments 1, among patients' baseline characters, only gender between disease groups is compared in the study. Did authors look into other baseline characters between CFS/ME and control group which may impact the diagnosis result? Good to discuss the finding on whether other baseline factors are balanced. 2, in the article, the less agreement between on individual physical signs (small Kappa coefficient) leads to a conclusion that they are not needed, which is not straightforward since the true outcomes of the physical signs is unknown. Further discussion on the interpretation of the result will be helpful.

REVIEWER	Robert M West University of Leeds, UK
REVIEW RETURNED	07-Jul-2017

GENERAL COMMENTS	This work is well presented in the manuscript. The statistical techniques were appropriate, the main statistical summaries according to standard practice and with suitable methods to calculate confidence intervals. The sample sizes have been checked and agreed. The 'gold standard' for determining caseness appears to be the usual UK NICE guidelines. The authors should confirm that this is so. The table of patient characteristics (Table 1) is extremely limited with only gender included. This is disappointing. Age of patients and controls would have been helpful. My main concern is that the study reports only 3 assessors(2 using the proposed technique), and that as a result inter-relater reliability is only assessed through kappa. This might be noted in the limitations of the study.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: B. Puri
Institution and Country: Imperial College London, UK Please state any competing interests: None declared

Please leave your comments for the authors below

Comment: The results of this relatively large and impressive study are well described. The authors correctly mention the fact that the two groups were not matched for sex; this does not affect the relevant values of Cohen's kappa. They rightly also draw attention to the need for further studies which examine more severely affected patients.

Response: Thank you for your comments.

Reviewer: 2
Reviewer Name: Dr Jo Daniels
Institution and Country: University of Bath Please state any competing interests: None declared

Please leave your comments for the authors below

Comment: This is a very interesting paper that is potentially of benefit to patients and clinicians, with the prospect of a more expedient and reliable method of diagnosis.

Clear introduction, well written. Perhaps underplaying the current utility of the Fukuda diagnostic criteria that is used across clinical services - an accepted method for diagnosis. Context is important.

Comment: I appreciate that a technique such as this would be very useful in clinical settings, however I am unconvinced by this paper. It is lacking in detail and without a comparison group and clearer explanations of methods and rationale for particular methods I do not feel it is of sufficiently publishable standard for the BMJ.

Response: We have added detail to the methods in particular around inclusion/exclusion criteria and recruitment.

Comment: I would suggest that the authors consult a statistician re potential ROC analysis - I am not adequately qualified to advise in this area, however it is worth recommending the consideration of further more sophisticated tests.

ROC analysis is another way of presenting the same sensitivity and specificity statistics as included in Tables 3 and 4. In this case, given that we are not considering a continuous test measure and have a low number of physical signs, a tabular form is a clearer and more complete way of presenting the relevant statistics.

Comment: For me the originality is there, but the scientific rigour is not, this may be addressed in a substantial revision to aid transparency and rigour. I thank the authors for their interesting contribution.

Response: We have made substantial revisions to the manuscript in particular the methods.

Specific comments:
1. Reference line 35, author pg 2 - Health and Care excellence? Not in refs. Do you mean to refer to NICE here?

Response: Corrected

2. I understand there is a protocol to refer to, but there is insufficient information here to make sense of the method of the study. Please say more.

Response: Corrected and we have also submitted the protocol as a supplement.

3. Please insert ethics numbers etc. in a more appropriate place, ethics section perhaps>

Response: Corrected, we have put this at the beginning of the methods section.

4. When you say 'circulated to the wider community' what do you mean? Please offer more detail in the procedure.

Response: We have removed this phrase and explained more clearly.

5. How did you recruit non-blood relatives exactly? The level of detail makes this nonreplicable, despite a protocol. Please make the protocol available as a supplement.

Response: This has been added and our protocol submitted as a supplement.

6. Inclusion/exclusion criteria please? I see you have mentioned this later, however a more traditional methods section with inclusion and exclusion would be more suitable.

Response: Now included

7. stating 'briefly' is feels inappropriate here; as scientific paper details are required for replicability. Expand: what are you not stating?

Response: Removed the term briefly as it was not appropriate.

8. What is the rationale for the 4+ inclusion and exclusion criteria? Is this evidence based?

Response: For the control group this was to assure that individuals were indeed healthy controls with no additional conditions with similar symptoms which could be confuse with CFS/ME. Individuals with CFS/ME were further screened with NICE (2007) and the international consensus criteria prior to inclusion.

9. Please describe/introduce the 'Perrin' technique before referring to it.

Response: This is now introduced/described earlier in the introduction section.

10. The authors report that there is no conversation between the participant and the examiner. Can you perhaps consider the vast literature relating to patient posture and facial expressions etc that may have unblind them?

Response: This is a really interesting point, however the illness behaviour assessor was no better than chance in his diagnoses of CFS/ME.

11. I'm unconvinced of the utility of this without a comparison group, i.e how would fibromyalgia patients fair here? Or patients with HIV?

Response: As it stands this was the first stage to see is there is any suggestion that the physical signs can aid diagnoses/screening as a proof of concept. A further study could indeed explore this further however as it stands this is beyond the scope of this current study. We accept this is a limitation and have expanded the “Limitations and suggestions for future research”.

12. What level of training did the clinicians have? AHP and others.

Response: Additional detail has been added.

13. The statistics appear to be fairly rudimentary here - would a ROC analysis not be appropriate?

Response: This would be an appropriately sophisticated analysis for reporting sensitivity and specificity of clinical test.
ROC analysis is another way of presenting the same sensitivity and specificity statistics as included in Tables 3 and 4. In this case, given that we are not considering a continuous test measure and have a low number of physical signs, a tabular form is a clearer and more complete way of presenting the relevant statistics.

14. Please correct the reference style so that it is in line with the journal guidance, e.g. how many authors listed in text and style.

Response: Corrected

15. Please be more transparent on the consent process, this is unclear.

Response: Corrected

Reviewer: 3
Reviewer Name: Jiawen Zhu, statistical scientist Institution and Country: Roche Innovation Center, USA Please state any competing interests: None declared

Please leave your comments for the authors below

Introduction
The authors present a study to examine the agreement of a physical assessment technique for CFS/ME. The goal of the study is to assess the 5 physical signs and comparing with the standard clinical neurological and rheumatologically examination. The authors conducted a diagnostic accuracy study with CFS/ME patients and a control group. Through sensitivity and specificity analyses of diagnosis results by three practitioners, examination of the diagnosis bias with McNemar's test and agreement between both diagnosis result and individual physical signs from experienced and newly trained AHP, the authors concluded using the physical signs appears to improve the accuracy of CFS/ME diagnosis results and only two of the physical signs may be needed based on the agreement assessment.

Major comments
1, by using one practitioner in each category, the study didn't consider the within category variation by practitioners, i.e. experienced AHP, newly trained AHP and physician. This is a major limitation to interpret the research result and make the conclusion to the general level, i.e. using the physical signs improves the accuracy of identifying patient with CFS/ME. A study design including diagnosis result from several experienced AHP, newly trained AHP and physician will help to answer the research question.

Response: We agree. As it stands this was the first stage to see is there is any suggestion that the physical signs can aid diagnoses/screening (i.e. to demonstrate 'proof-of-concept' of the Perrin technique). A further study could indeed explore this further however as it stands this is beyond the scope of this current study.

Minor comments
1, among patients' baseline characters, only gender between disease groups is compared in the study. Did authors look into other baseline characters between CFS/ME and control group which may impact the diagnosis result? Good to discuss the finding on whether other baseline factors are balanced.

Response: Unfortunately, no other baseline characteristics were recorded. We have included this as a limitation in our discussion section.

2, in the article, the less agreement between on individual physical signs (small Kappa coefficient) leads to a conclusion that they are not needed, which is not straightforward since the true outcomes of the physical signs is unknown. Further discussion on the interpretation of the result will be helpful.

Response: We have added to the discussion “However, this does not necessarily mean that the signs were not present; it could mean that the newly trained AHP found these signs more difficult to detect. Despite this, there was moderate agreement between both AHPs on overall diagnosis.”

Reviewer: 4
Reviewer Name: Robert M West
Institution and Country: University of Leeds, UK Please state any competing interests: None declared

Please leave your comments for the authors below

This work is well presented in the manuscript.
The statistical techniques were appropriate, the main statistical summaries according to standard practice and with suitable methods to calculate confidence intervals. The sample sizes have been checked and agreed.

Comment: The 'gold standard' for determining caseness appears to be the usual UK NICE guidelines. The authors should confirm that this is so.

Response: This is now added

Comment: The table of patient characteristics (Table 1) is extremely limited with only gender included. This is disappointing. Age of patients and controls would have been helpful.

Response: Unfortunately age was not recorded. We have included this as a limitation in our discussion section.

Comment: My main concern is that the study reports only 3 assessors (2 using the proposed technique), and that as a result inter-relater reliability is only assessed through kappa. This might be noted in the limitations of the study.

Response: As it stands this was the first stage to see is there is any suggestion that the physical signs can aid diagnoses/screening. A further study could indeed explore this further with more assessors, however as it stands this is beyond the scope of this current study.

VERSION 2 – REVIEW

REVIEWER	Jiawen Zhu Roche Innovation Center, U.S.A
REVIEW RETURNED	03-Sep-2017

GENERAL COMMENTS	Thanks for addressing the previous questions, no more comments.