Author's response to reviews

Title: Agreement between a self-administered questionnaire on musculoskeletal disorders of the neck-shoulder region and a physical examination

Authors:

Nathalie Perreault (nathperro@hotmail.com)
Chantal Brisson (chantal.brisson@uresp.ulaval.ca)
Clermont E. Dionne (clermont.dionne@uresp.ulaval.ca)
Sylvie Montreuil (Sylvie.Montreuil@fss.ulaval.ca)
Laura Punnett (Laura_Punnett@uml.edu)

Version: 3 Date: 17 October 2007

Author's response to reviews:

October 17, 2007

Dr. Deborah Saltman
Editorial Director (Medicine)
BioMed Central Ltd
Science Navigation Group
Middlesex House
34-42 Cleveland Street
London W1T 4LB
United Kingdom

Object: Submission of a revised manuscript to BMC Musculoskeletal Disorders (1605808436128597)

Dear Dr. Saltman,

Enclosed is a revised version of the manuscript titled "Agreement between a self-administered questionnaire on musculoskeletal disorders of the neck-shoulder region and a physical evaluation" we would like to submit for consideration for publication in BMC Musculoskeletal Disorders. This paper reports the results of an original study conducted among video display unit (VDU) users that suggest that the agreement between self-reported symptoms and a physical examination of the neck-shoulder region is only fair to good, and is affected by the consideration of pain. We believe this topic is highly pertinent for the readers of BMC Musculoskeletal Disorders because of the burden of conducting individual physical evaluations in large study populations and the consequent frequent use of questionnaires to measure the prevalence of musculoskeletal disorders.
As requested, you will find in the next pages, the details of our response to each one of the reviewer's comments.

This manuscript reports original work that has not been published nor submitted for publication in other journals.

Thank you for your attention on this submission.

Sincerely,

Clermont E. Dionne
Corresponding author

RESPONSES TO THE REVIEWER’S COMMENTS

We thank the reviewer for his thoughtful comments.

Major Compulsory Revisions

Comment #1: “The phenomenon of (questionnaire defined) symptoms versus clinical signs should be discussed. Are they measuring the same concept?”

As requested, we have discussed this issue on page 16, line 16: “The definition based on questionnaire may not measure the same concept than the physical examination. While the physical examination measures the integrity and the absolute performance of the structures and tissues, self-reported symptoms are based on actual performance and sensation, much affected by pain perception. This distinction is supported by the large impact that pain has on the agreement. The results of this study suggest that pain intensity is an important feature in the agreement between a questionnaire on musculoskeletal disorders and a physical examination and support the construct validity of a case definition based on symptoms.”

We have also added a sentence in the abstract: “a fair to good agreement between the presence of musculoskeletal disorders ascertained by self-administered questionnaire and physical examination that may reflect important differences in the constructs measured. Investigators should consider these results before choosing a method to measure the presence of musculoskeletal disorders in the neck-shoulder region.” And in the conclusion on page 18, line 8: “Investigators should consider these results before choosing a method to measure the presence of musculoskeletal disorders of the neck-shoulder region.”

Comment #2: “The definition of non-cases leads to the possibility that they might as well be quite unhealthy, e.g. symptoms approaching 100 on the VAS-scale during two days last week or pain 40 mm every day last week. This leads to a misclassification of caseness and tends to attenuate possible true associations
with e.g. medical examination results. This should be mentioned in the discussion.

As requested, we’ve added a paragraph on page 17, line 1 to discuss this issue:

The higher prevalence of findings in the physical examination than in questionnaire might be due to the selection criteria used to define non-cases according to the questionnaire. Given that the questionnaire definition was somewhat restrictive, some non-cases were not totally free of symptoms. Indeed, 26 of those 102 workers classified as non-cases according to the primary questionnaire definition had symptoms in the week prior to the questionnaire. This could have lead to a classification bias and could have attenuated the true associations with physical examination.

Minor Essential Revisions

Comment #3:

a) How about generalizability of the results when about 65% of all subjects were female clerical workers, and response rate somewhat low among non-cases. Very few men.

As suggested, we’ve added a sentence about the generalizability of the results on page 17, line 21 in the discussion: Finally, the current study population consisted mainly of employed clerical women, thus the generalizability of the results is limited to similar populations.

b) Table II is of minor relevance for this study. It tells that the result probably may be generalized to the 627 VDU-users at this organization, but nothing about others.

We’ve included the information of Table II in the first paragraph of the results section (page 9, line 17) and then, removed Table II: The VDU users in the agreement study were similar on demographic and occupational characteristics to all VDU users. Study participants were primarily female (83%). The mean age was 44 years. More than 80% of the participants were clerical workers, 11% were professional and executives and 7% were technicians. The average use of VDU was 20 hours per week.

We’ve modified the numbers of the subsequent tables accordingly.

Comment #4:

a) Page 7 bottom - it is unclear how many - one or two examined the subjects.

As suggested, we’ve clarified the sentence, page 7, line 8: After standardization of the procedure, the same trained occupational therapist, blind to the participants’ questionnaire answers, performed all the physical examinations.
b) Page 8 - muscle strength testing was based on manual testing. The reliability and validity is questionable. This should be discussed and could theoretically explain some of the lack of association with symptoms.

As requested, we’ve added elements to discuss this issue on page 13, line 22:

According to previous studies, tests used in physical examination, especially measurement of range of motion and manual muscle testing, have poor to good reliability [47-53]. Also on page 14, line 5: Nevertheless, the comparisons in the present study might have been compromised at least in part by measurement error which could explain some lack of association with symptoms.

Comment #5: Page 12 - what test of statistical significance? No such is described in the methods section. Overlapping CI of two point estimates does not mean that the CI of the sampling distribution of the differences between the point estimates includes zero, i.e. non-significance. Same comment to footnote on Table VI.

As suggested, we’ve added a sentence describing the statistical test performed in the data analyses paragraph on page 9, line 10: Finally, a stratified analysis was performed to determine the effect on the agreement of the time elapsed between the administrations of the questionnaire and the physical examination. The chi-square test was used to compare percentages.

We’ve also modified the sentences in the results section, page 11, line 21: The highest Kappa value of the study (k=0.54) was obtained when the questionnaire and the physical examination were administered 21 days or less apart. The better agreement observed with the shortest period (21 days or less) between the administrations of the two tests was reflected in both cases and non-cases, however, none of the comparisons reached statistical significance because of the limited sample size (p-values were respectively 0.10 for global agreement, 0.30 for agreement among cases and 0.31 for agreement among non-cases).

The footnote on Table VI was also modified accordingly. None of the comparisons reached statistical significance (χ² test). (Note that Table VI is labelled Table 5 in the revised version of the manuscript.)

Discretionary Revisions

Comment #6: Likewise relevant would be to evaluate the influence of non-case definition. One most relevant alternative definition would have been symptom-free subjects. These would be the tentative counterparts to subjects without clinical signs.

In supplementary analyses, 26 subjects with symptoms who were classified as non-cases (because they did not meet the inclusion criteria of cases for symptoms duration and/or intensity) were moved to the group of cases and measures of agreement were estimated again with the new groups of subjects. This comparison (new questionnaire case definition vs primary definition by physical exam) yielded the following estimates: Kappa=0.47 (0.34-0.59), global
agreement=74%, agreement among cases=74%, agreement among non-cases=74%, sensitivity=80%, specificity=66%. Hence, with this approach sensitivity increased but specificity decreased compared with the original definition (at least 3 days of pain with an intensity of at least 50/100 mm). Also, global agreement had changed little (74 vs 72%).

Comment #7:

a) ¿The data could also be analysed calculating sensitivity and specificity.¿

As suggested, we performed analyses calculating sensitivity and specificity. The values of sensitivity and specificity have been added in the Tables. Also, we’ve added a sentence in the data analyses paragraph on page 9, line 8: ¿Percent agreement among cases and non-cases, which corresponds to positive and negative predictive values, were also calculated, as well as sensitivity and specificity [43].¿

We’ve described the results of these analyses in the manuscript:

- Page 10, line 22 (results of Table 3): ¿The inclusion of the functional criteria to the primary questionnaire definition increased specificity but decreased sensitivity figures.¿

- Page 11, line 1 (results of Table 4): ¿When the primary questionnaire definition was compared with the three physical examination definitions, the Kappa varied from 0.30 to 0.48 (Table 4). The Kappa was lowest (0.30) when the physical definition was based only on decreased range of motion or muscular strength. The global percent agreement (66%), sensitivity (64%) and specificity (67%) were also somewhat lower with this definition.¿

- Page 12, line 5 (results of Table 5): ¿A higher sensitivity was also observed when the questionnaire and the physical examination were administered within 21 days (sensitivity = 75%) than over 21 days (sensitivity = 56%).¿

b) ¿The agreement among cases and non-cases are traditionally called ¿positive¿ and ¿negative predictive value¿ respectively. This could perhaps be mentioned.¿

As suggested, we’ve added this terminology in the corresponding sentences of the methods and results sections, and in the corresponding Tables. Methods, page 9, line 8: ¿Percent agreement among cases and non-cases, which corresponds to positive and negative predictive values.¿ Results, page 10, line 19: ¿Percent agreement among cases (positive predictive value) tended to increase with the inclusion of functional limitation criteria (Table 3). For the non-cases, global percent agreement (negative predictive value) varied little, remaining around 60% for all functional limitation definitions.¿

Comment #8: Page 8 ¿Please explain why two types of VAS scales were used 1-100 mm and a 11 point scale. Is the latter a true VAS scale, or just an ordinal scale?¿
A 100 mm-VAS was used in the questionnaire and an 11-point numerical rating scale (NRS) was used during the examination. NRS are more appropriate to use in face-to-face and telephone interviews than VAS, and their psychometric qualities are comparable to those of VAS. This has been clarified on page 7, line 23: “After performing each maneuver, the subject was asked to record his/her pain level on an 11-point numerical rating scale (NRS). NRS are more appropriate to use in face-to-face and telephone interviews than VAS, and their psychometric qualities are comparable to those of VAS [35, 36]. Pain was considered significant if it was reported at the relevant site during maneuvers and was of moderate intensity or worse (a score of 3/10 or more).”

Comment #9: “Page 14, 2nd paragraph about recall bias could perhaps be relocated to the paragraph on previous page discussing reliability and validity issues.”

As suggested, we relocated this paragraph on the previous page (page 13, line 5) discussing validity issues: “Previous studies also provide evidence of construct validity of subjective symptoms reported in questionnaires [46]. Also, VAS are considered among the best instruments to measure pain [32]. To reduce the impact of potential error in recall in this study [41], only symptoms in the last seven days were considered.”

Comment #10: Page 16, 2nd paragraph “Statement about the consistency of the results with the natural course of musculoskeletal disorders is relevant if questionnaire based pain is related to pain provoked by the examination manoeuvres. Sometimes yes, sometimes no.”

This sentence has been reformulated (page 16, line 9): “These results are consistent with the hypothesis that musculoskeletal disorders are progressive and that patients may have symptoms before objective physical findings appear [58].”