Reviewer's report

Title: Prediction of trapezius muscle activity and shoulder, head, neck, and torso postures during computer use: results of a field study.

Version: 3 Date: 13 March 2014

Reviewer: Andrew Judge

Reviewer's report:

The aim of this study is to describe the effect of computer use on development of musculoskeletal symptoms.

The authors say how exposure to computer use could be measured using two methods:
1) a “task based” activity keyboard/mouse/idle activities only
2) a “full model” which is “task based” AND, questionnaire, workstation, and anthropometric parameters

The outcomes of interest chosen to represent musculoskeletal symptoms were:
1) range of amplitude
2) trapezius muscle activity

Within the paper a number of metrics are used to assess the performance of these different models on predicting these outcomes, including R-squared (how well the model explains variation in outcome) and RMS (residual error between model predicted values versus observed data values). Essentially, is a simple task based model adequate, or is a more complicated model including more detailed workstation assessment required?

From the methods section it becomes clear that each individual in the study has for each outcome (all measured trapezius muscle activity and shoulder, head, neck, and torso postures) measured outcome as, i) their median value, and ii) the range being the difference between the 10th and 90th centile.

Comments:
- I'm not keen on the wording of “full model” and “task based”. Its essentially a simple versus detailed/comprehensive measure of exposure. Also full model implies a saturated model with all 104 variables, but it is actually a reduced model from a model selection process that contains only the significant predictors of outcomes.
- The sample size is only 117 people, with 104 measures in the full model? If this represents 104 degrees of freedom then the model is overfitted? As a rule (see Harrell F book) there is a rule of 10 to 1. So for a continuous outcome, if there are 117 people, then maximum of 17 degrees of freedom (not necessarily the same as variables) in the model.
Whilst I’m aware the authors have used some form of model selection techniques, such as categorizing the 104 variables into seven smaller subgroups of variables, and doing backward selection in each subgroup, there is still the potential for overfitting. Towards the end of the discussion, there is some mention of this as a limitation (issues of multiple testing and possible chance findings (type I error)), but this needs expanding.

- It should be acknowledged as a limitation that some of the significant predictors identified in the full model, may be chance findings that are significant in this sample of 117, and not necessarily replicated in other studies or samples.

- Could the authors justify their approach to model selection? Did you consider using approaches such as bootstrapping to get a bias corrected measured of R-squared to get a feel for the performance of the model in other samples of data?

- For the regression models, were the assumptions satisfied e.g. normality of residuals, etc. Were the distributions of the outcome measures normally distributed.

- Previous comments that the more complicated model would obviously have better model performance is not necessarily true. It is not the full saturated model, and instead a model of significant predictors after model selection. Whilst I agree that you can improve R-squared by adding in lots of variables to a model that are not significant and it overfits the data to the sample. I don’t think that is the case with this study.

- What is clear is that the more complex model is much better at explaining the outcomes.

Declaration of competing interests:

I declare I have no conflict of interests