Author’s response to reviews

Title: Prognostic factors of a favorable outcome following a supervised exercise program for soldiers with sub-acute and chronic low back pain

Authors:
Marc Perron (marc.perron@fmed.ulaval.ca)
Chantal Gendron (chantal.gendron@forces.gc.ca)
Pierre Langevin (pierre.langevin@fmed.ulaval.ca)
Jean Leblond (jean.leblond@cirris.ulaval.ca)
Marianne Roos (marianne.roos.1@ulaval.ca)
Jean-Sébastien Roy (jean-sebastien.roy@fmed.ulaval.ca)

Version: 2 Date: 31 Jan 2018

Author’s response to reviews:

Mark Hancock (Reviewer 1): The authors have taken on most of my previous suggestions and I believe the manuscript is substantially improved. Given the major changes requested by myself and the other reviewers I have a few further comments on the revised manuscript, mostly relating to the analysis/modelling.

1. The model building process is still unusual and requires further justification/clarity. I understand the desire to not be held to statistical significance in this hypothesis testing study, but the criteria for retaining variables is still unclear. Could the authors base the decision to retain variables in multivariable model on the strength of association (i.e OR). Maybe OR greater than 2 even if not statistically significant. The methods should be modified to provide more clarity.

As suggested by the Reviewer, the variables were retained when the Odds Ratios were above 2, even if the p-value was above 0.05. The following was added to the Results section (Lines 300-
Furthermore, as the five variables had an odds ratio above 2.0 in this final regression, we decided to retain all variables in the final model.”

Given the final model includes variables that are not statistically significant, remove/modify statements that refer to identifying 5 prognostic factors, especially in key sections like abstract, first sentence of discussion and conclusion. Consider language something like, "The final model included 5 variables…."

The wording was changed as suggested by the Reviewer.

Related to points above the manuscript would be clearer to me if the word multivariable was used consistently in methods and results (and tables) do differentiate from the univariate analyses.

Done as suggested

In methods paragraph starting on line 219, can the word collinearity be included as this term would be more familiar to most readers and I think this is the key element here.

As suggested by the reviewer, we included the word collinearity and reworded the sentence to make the procedure clearer for the reader. The sentence has been modified as follows: “The KMO index and measures of sampling adequacy (MSA) were used to determine collinearity. When the KMO index is greater or equal to 0.6, the logistic regression may be performed with all the potential prognostic factors. When the MSA are below 0.6, the associated potential prognostic factor must be removed.” (Lines 220-223)
5. Line 230-231: avoid the terms "carried both common and unique information", and instead provide clear information of the criteria used when deciding whether variables remained or not in the multivariable model (consider the OR suggestion above, or provide some other criteria).

The terms “carried both common and unique information” were changed for “had p-values below 0.50.” (Lines 230)

6. Line 288: similarly, to above avoid "had a clear contribution" and use statistically significant which is much clearer. There are other parts of manuscript where similar changes would enhance clarity.

Changes made as suggested. For example, “had a clear contribution” was changed for “had p-values below .05” (lines 284 and 304)

7. Make it clearer which model the tables refer to.

The models were added in the Tables

8. Line 384: remove likely to benefit as the authors have done in the rest of the manuscript.

We modified the sentence as follows : “The present study established five variables to identify patients most likely to have a favorable outcome, regardless of their participation in the exercise program.” (lines 381-382). We also modified the last sentence of the conclusion as follows : “Future validation studies should be carried out with other populations to confirm this CPR and
subsequently, to verify whether some of these factors may be considered treatment effect modifiers.” (lines 383-386). We think that this emphasizes the fact that further studies are needed before considering these variables as treatment effect modifiers.

Paul Bruno (Reviewer 3): The authors have generally addressed my initial comments. Two additional comments for the authors to consider are:

1. The new stated sample size requirement indicates that 90 participants were required. However, only 85 participants completed the study. This should be highlighted in the limitations section.

As suggested, a limitation related to sample size was added: “Finally, the targeted sample size of 90 participants to be included in the statistical analyses was not met, as only 85 of the 104 participants took part in both evaluation sessions.” (Lines 376-378)

2. Although there was no between-group difference in the number of sessions attended, it would be helpful to add a suggestion to future researchers in this area that the exercise "dose" be standardized.

Although this is an excellent suggestion, there is no section in the discussion related to the adherence, as the adherence was excellent in the present study. Therefore, it would be a stand-alone sentence in the discussion. Furthermore, it is very difficult to have a standardized ‘dose” for a population such as soldiers.

Corey Simon (Reviewer 4): The authors have attempted to address most of my concerns, and have scaled back the interpretations of this study. In particular, the revised study aim is to
explore potential predictors of the multi-stage exercise program as a means to inform the next iteration of hypothesis-testing trials - and potential development of a CPR. While I consider the revised manuscript improved, I have a few remaining concerns:

1. Given the unavoidable flaws of small sample size, model over-fitting, and non-significant factors in the final model, I'm not sure there is value in attempting to determine predictive capacity by number of criteria. As you have already stated, you're at the stage of generating hypotheses. You've achieved this by looking a correlations with favorable vs. unfavorable response for this intervention. In a subsequent study (with larger sample size and factors determined a priori), you can then consider predictive capacity based on number of criteria.

Thank you for this comment and we fully understand the relevance of it. However, we feel that we need to give clinicians an idea of the predictive capacity of the retained variables. In the discussion, however, we clearly state: “Thus, before continuing to the validation step, our results need to be confirmed in a hypothesis-testing study in which a limited number of a priori hypotheses will be tested and appropriate adjustments for multiple comparisons will be made.”

In the conclusion, we also state that: “Careful use of these variables is mandatory for clinical purposes as this study is at the early stage of CPR development.”

Therefore, while we attempted to determine predictive capacity by number of criteria to give clinicians an idea of the value of the present results, we also warn them that our results still need to be validated in a larger sample.

2. Regarding my initial concern of a subgroup of individuals with acute LBP receiving PT before the program, the authors note that: 1) baseline variables were collected after the participants had
received PT for acute LBP, so prior treatments should not affect the present results; and 2) program was not designed for acute LBP. However, this doesn't change the potential for different subgroups in your study based on pain duration and last medical intervention. This is less of a problem if the baseline measures are compared between the group with acute LBP coming off PT (assuming they're now considered subacute) and the rest of the cohort, and found to be similar. Nowhere do I see that comparison, or a test to determine if either group had a propensity for favorable vs. unfavorable outcome. What the baseline measures won't inform is the extent to which time since medical intervention (and potentially response to that intervention) influence participant expectation of the supervised exercise program - which can affect outcome.

We thank the reviewer for this comment. Looking at Tables 1 and 2, one can see that subjects in both subgroups received treatments before the initial evaluation (mean of 3.14 vs. 6.85). Therefore, most of the included soldiers received treatments before the program. Furthermore, even though the Unfavorable outcome subgroup received significantly more treatment before the initial evaluation than the Favorable outcome subgroup, the score on the Oswestry Disability Index and the mean LBP perceived in the last 48 hours is similar in both groups. The program was developed for subacute and chronic LBP. The goal of the study was to identify variables of a favorable outcome in soldiers with sub-acute and chronic LBP participating in a multi-station full-body supervised exercise program. Therefore, we don’t see the value of comparing the baseline measures between the group with acute LBP coming off PT and the rest of the cohort as we want to know the variables associated with a favorable outcome once they are not in their acute phase. We have the data to perform the analysis and we could add this analysis if the editor feels that we should, however we don’t see the added value of this analysis.

3. Baseline characteristics indicated that the average length of time since the last bout of LBP was 16 months. Does this mean there are individuals that were in the program who had not experienced LBP for months and arrived pain free? Or, is it supposed to indicate since latest onset of LBP? Assuming the latter, but please clarify.
As suggested by the reviewer, we replaced the term «episode» in the table 1 by «onset» to clarify that this variable indicates the time elapsed since the moment the pain had begun or increased. Note that in the methods section (lines 85-86), we specified that «the participants had to ...present with an episode of subacute or chronic LBP...» which indicate that the participants were symptomatic.

4. Since this study cannot generalize to non-military personnel, I think it's important for the reader to know as much. Please consider aligning your title with the population (e.g. soldiers or military vs. individuals). Also, your abstract conclusion still hints at guiding treatment of non-military personnel. There is nothing wrong with keeping the focus on military. In fact, it would strengthen the manuscript to discuss similar prognostic trials. One example is the POLM trial by Childs et al. (2011, 2014).

We agree with this suggestion. We replaced the term «individuals» in the title by «soldiers» and the term «patients» in the abstract by «soldiers with LBP» (line 25).