Author's response to reviews

Title: Construct validity of functional capacity tests in healthy workers

Authors:

Sandra E Lakke (a.e.jorna-lakke@pl.hanze.nl)
Remko Soer (R.Soer@cvr.umcg.nl)
Jan HB Geertzen (j.h.b.geertzen@umcg.nl)
Harriët Wittink (harriet.wittink@hu.nl)
Rob KW Douma (k.w.douma@pl.hanze.nl)
Cees P Van der Schans (c.p.van.der.schans@pl.hanze.nl)
Michiel F Reneman (m.f.reneman@cvr.umcg.nl)

Version: 2 Date: 3 April 2013

Author's response to reviews: see over
Dear Dr. Staal,

Thank you for critically reviewing our manuscript and for the constructive comments. We have carefully considered reviewer’s comments and made revisions accordingly.

Please refer to our comments and revisions listed below (in italics). Changes in the manuscript are indicated using red type. With these revisions, we feel that the quality of the paper has increased.

For the preparation of the second draft, we were assisted by a bio-statistician of the University of Groningen and the manuscript has been edited by a native-English speaker with scientific expertise.

We hope that you are satisfied with the revisions and that you are able to accept this paper for publication.

Sincerely,

On behalf of all authors,

Sandra Lakke

E-mail: a.e.jorna-lakke@pl.hanze.nl
List of changes for the revision of our manuscript: “Construct validity of functional capacity tests in healthy workers”

Response to reviewer #1: Jan Kool

Major Compulsory Revisions

1. External validity of results: In general, this cross-sectional research in healthy workers is important and the data basis (number of subjects studied and variables) is very rich. The result that non-physical factors do not concurrently explain FC is very surprising. My main concern is how these results should be interpreted. Are they valid for prognosis of future work disability? Probably not, this would have needed a longitudinal study. Are they valid for patients? This should also be discussed and included in the conclusion.

   We thank the reviewer for his help in stressing the importance of explaining the relevance of this study. We addressed the question by changing the sequence of sentences in the introduction section and adding two sentences (page 6, 2nd para): “FCEs facilitate the reasoning process for clinicians and assist them in determining if further examination is required [1]. FCEs also assist clinicians in pre-employment screening for healthy workers.” And (page 7, 2nd para): “From a clinician’s perspective, in healthy workers during pre-employment screening, knowledge of related factors is necessary to identify the necessity of additional testing.”

   In the discussion section we added a limitation and recommendation section to address the question of how these results should be interpreted (page 19 and page 20).

   In the conclusion section we added (page 21, 1st para): “Because of the cross sectional design and the healthy study sample in this study, the results should not be interpreted as predictors for future work performance, nor should they be generalized to patients.”

2. Definition of construct validity is not appropriate (background, third paragraph): reference [14], which is used for the definition, investigates test retest reliability and does not cover construct validity. The quoted article does not give a definition or description of construct validity. Generally, the quoted definition is questionable and should be reconsidered. Most readers may think about construct validity as it was defined by Portney and Watkins. According to these authors, construct validity reflects the ability of an instrument to measure an abstract concept, or construct [Portney & Watkins. Foundations of clinical research: applications to practice – 3rd ed. 2009].

   We have changed the text accordingly (page 7, 2nd para): “Construct validity is ... workers [14].”

3. Type of validity: The use of the term ‘construct validity’ needs clarification. Which are the consequences of your results? Would you suggest not assessing other than physical factors? That might save time. But this cross sectional study is not suitable to determine predictive validity. A longitudinal study would be needed to determine whether non-physical factors assessed in pre-work screening are predictive for future work disability. And if this would be the case, not assessing these is not justified.

   We agree. We followed your advice and added a limitation section of our findings in which we describe that we did not study predictive validity for future work disability (see comment 1). We added a recommendation section (see comment 1).
4. Choice of FCE variables is unclear. Why are these FCE tests chosen? Do they cover FC? This could be explained in the background or discussed in the discussion.

   *Thank you for your attention to this. The tests in this study are operationalizations of work-related FC. We added the words “work-related” to FC in the Abstract, Background and Discussion section (Page 6, 2nd and 3rd para and page 20, 2nd para).*

   *These FCE tests are chosen to cover a range of physical activities. The ICF activities ‘carrying, moving and handling objects’ (ICF d430 and d449) and ‘maintaining a body position’ (d415) were covered [1].*  

   *In the method section, we added a sentence to better address the choice of the tests (page 9, 1st para): “Functional capacity was measured with five FCE tests, selected to cover a range of physical activities.”*

5. Reporting and description of variables and methodology is confusing and inconsistent: the paragraph ‘construct validity’ should be revised because of missing punctuations. Table 1 is unclear, because the numbers in column ‘factor’ are not related with the manuscript and table 3. Generally, description and reporting of variables should be consistent through the whole manuscript, figures and tables.

   *To address this point, we edited our manuscript by a native-English speaker with scientific expertise. We replaced 18 for 15 at the end of the Background section. We checked the description and tables again for consistency.*

6. Table 1: Please place headers of the 2nd column consistently in an ‘empty line’ as you did with ‘Personal factors’. I don’t understand why ‘education’ is a physical factor. This seems not plausible to me and may not be in accordance with the ICF.

   *We changed Table 1 accordingly.*

   *We addressed this comment and now we used the ICF linking rules according to the way other ICF experts linked this factor [1-6]. We linked “Education” to the ICF Participation component (ICF d820).*

7. Table 2 and limitations of a cross sectional design: non-physical factors did not explain FC. This is very surprising and I think, in the context of so many other findings, needs careful consideration in this publication. Do you think that the (limited) range on non-physical variables may have contributed to not finding associations? If there is insufficient variance in e.g. ‘perception of work’ factors, then associations cannot be detected. In addition, do you think that a longitudinal design identifying patients who develop work related disability could identify some factors as predictive for future work disability? This might be valuable to discuss. Abbreviations r, p and rpbi should be clarified.

   *Functional Capacity is a multidimensional construct (page 6, 1st para). The results of our study indicate that the balance between these dimensions differs between healthy workers and patients groups. Most other FCE validity studies have been performed on patients (with pain), and these studies did indeed demonstrate relationships between psychological factors and FC. This might have led to the implicit assumption that these relationships should be similar in healthy workers, and consequently the surprise to find out that we did not find these results. We feel that we have carefully executed this study, the analyses, and interpretations, and*
concluded that these relationships appear different between patients and healthy workers. We have not been presented with compelling arguments from the reviewer(s) that this study was incorrectly executed or interpreted. However, this study would definitely benefit from replication in a different sample of healthy workers to analyze the robustness of our results.

Page 20, 2nd para: "We recommend researchers to replicate this study in a different sample of healthy workers to analyze the robustness of current observations."

The reviewer suggests that different or broader operationalizations of the constructs might have led to different results. Other currently not measured physical and non-physical variables might have contributed to not finding associations. We addressed that in the discussion section and rewrote it to provide more clarity (page 18, 2nd para to page 19): “Other biopsychosocial factors.....endurance [63].”

We checked the range as defined as the difference between the lowest and highest values, and we checked percentiles with the statistician. The variances in perception of work were broad and, if a ceiling effect was present, we corrected it (page 13,3rd para). The variance in the psychological factor ‘mental health’ was small (Table 2). Based on these findings, we concluded that the amount of variance could possibly explain the absence of low correlation between mental health and FC tests (page 18, 1st para): “However, an explanation for finding no associations between FC tests and mental health scale in our study might be, beside the absence of chronic pain, that the small variance encountered of the mental health scale may explain the current results (Table 2).”

We do think that a longitudinal cohort study of healthy pre-screened workers might identify risk factors for future work disability [7]. We added (page 19, 2nd para): “The cross sectional design is not suitable for prediction of future work performance or future work disability. Therefore no conclusions to bio-psychosocial factors that may possibly be influencing future work performance or work disability can be made based on this cross section study.”

Abbreviations: The change is implemented in the method section and Table 1 (page 13, 3rd para):" The value of Pearson (r), Spearman (ρ), and point-biserial correlations(r_{pb}) were interpreted as being strong for significant (P_{bonf} < .002) correlations when r, ρ, r_{pb} > 0.75; moderate when 0.50 < r, ρ, r_{pb} ≤ 0.75; fair when 0.25 < r, ρ, r_{pb} ≤ 0.50; and low when r, ρ, r_{pb} ≤ 0.25 [14]."

8. Table 3: correlations, e.g. of muscle power, are much lower for different genders then over all. Could you explain why? The reason may be that women and men differ in weight and height. Why didn’t you use the gender specific correlations? Why didn’t you correct for body height and weight (partial correlations)? There probably is a good reason and I think it is important to explain this procedure. Classification or description of top line should be corrected, because correlation testing does not require dependent and independent variables. Could you explain why? Why didn’t you use the gender specific correlations? Why didn’t you correct for body height and body weight (partial correlations)? Classification or description of top line should be corrected, because correlation testing does not require dependent and independent variables.

As you can see in the scatter plots below and in Appendix 2, there is indeed a correlation between gender, muscle power, height and weight. Gender is the strongest predictor for FC lifting low. The correlation is lower for the different genders because both (strength and gender) reinforce the correlation on the same positive line. We added an explanation of this into the result section (page 15, 1st para): “Gender is fairly correlated to material handling FC tests and strongly correlated with handgrip strength. Elbow flexion inter-correlated significantly and strong with elbow extension, shoulder abduction, and handgrip strength (Appendix 2).”
We added Appendix 2 to provide information about inter-correlations.

For the preparation of the second draft, we were assisted by Prof. Dr. W.P. Krijnen, biostatistician of the Hanze University of Groningen. He advised us to use the minimum BIC criterion. By minimum BIC we dealt with the possibility for each FC test that weight or height counts or did not account for extra explainable variance, or even could replace gender (as suggested). The reported findings reveal how parsimonious models deal with this issue. In some FC tests gender could not be replaced by strength, height or weight (lifting low and carrying) and in others (standing forward bend, overhead working) gender could be replaced by e.g. weight or strength. The results indicate that a specific view for each FC test variable seems best to provide insight in whether strength, height, and weight are useful predictors in additional to gender.

We removed the words dependent and independent variables in Table 3, in the method and result section of the abstract as well as in the results section of the manuscript.

9. Motivation for the choice of potentially explaining variables is unclear. Why was back muscle strength not assessed? The fact that prolonged forward bending and overhead work could not be explained by physical factors raises the question whether other physical tests (aerobic static submaximal) might be more appropriate than strength.

   We agree and added a sentence in the discussion section (page19, 1st para): “Finally, in regard to the domain body functions, muscle endurance was not measured in this study and may correlate with static work FC tests, especially low back muscle endurance [63].”

10. Healthy persons were evaluated in this cross sectional study. Physical factors explain FC. These results may contribute to pre work screening in healthy persons. Can these results be applied to patients? On p18, at the end of the 2nd par, you mention the ‘incongruence of the profile (please define what is meant by this) between healthy workers and patients. Therefore, could you please discuss whether you think the results from this study are also applicable to patients? This is
probably not the case as non-physical factors appear to be more important in patients. So what I think is very important for this publication and its future interpretation in the work rehabilitation field is, whether you think the construct of FC is the same in healthy workers and in patients. If this is not clarified there is a risk of unintended misinterpretations by readers. I think an according remark that these results are applicable for pre work screening in healthy workers, but not for patients, should be included in the conclusion.

We agree that these results cannot be applied to patients and added this comment in the Discussion and Conclusion section (page 19, 2nd para): “The cross sectional design is not suitable for prediction of future work performance or future work disability. Therefore no conclusions to bio-psychosocial factors that may possibly be influencing future work performance or work disability can be made based on this cross section study.” And page 21, 1st para: “Because of the cross sectional design and the healthy study sample in this study, the results should not be interpreted as predictors for future work performance, nor should they be generalized to patients.”

We have changed our word choice (incongruence in the profile) in the introduction (page 7, 2nd para) and discussed the differences between the current data from a sample of healthy workers and data from previous studies (page 17, 4th para and page 18).

The FC construct is multidimensional. Based on the results of this study, compared to previous studies in patients, there are differences in the balance between the ICF dimensions as they relate to FC test results between healthy workers and workers with health conditions.

11. Discrepancies between your study and previous research: Some references in the background mention the importance of non-physical factors. These studies may have used different designs and sampling then your study. Could you please discuss possible reasons for the discrepancies between your findings and those of other studies, e.g. those in the background?

In this second draft, discrepancies are discussed more extensively in the discussion section (page 17, 4th para, page 18, and page 19).

Minor Essential Revisions

12. Abstract should be fully revised: first sentence is incomplete. The methods should state what the variables of functional capacity and other ICF factors are evaluated.

We changed the abstract according to this comment and added the variables of functional capacity to the method section of the abstract (page 4).”

13. Background does not underpin the research hypothesis: the authors mention the assumed correlation between the different ICF components but do not explain why these associations are questionable. Please explain how the hypothesized low to fair relationships (Table 1) were defined.

The associations are questionable because they are based on untested assumptions. To clarify this in the text, we re-arranged the background section (page 7, 1st para): “Until now, the assumed relationships have not been tested in healthy persons. It is of importance to analyze these assumed relationships in a sample of healthy workers, because it helps to understand what we are actually testing [13], which is important theoretically to unravel the construct of FC and to develop valid FC tests for healthy workers.”

We added the explanation of low and fair correlation in the legends of table 1 (page 22).
14. p6 2nd para 1st sentence: you may want to limit this statement to healthy persons. As this study and comparison with other research shows, the FC may be determined by factors that are very different in healthy persons and patients.

*We rewrote the construct validity paragraph (page 7, 2nd para): “From a researcher’s perspective, a comprehensive set of factors related to FC test results in healthy workers may perform as a reference to compare patients’ relationships between FC tests and ICF factors.”*

15. P9 in the middle: ‘In this method …[18,19]’: This sentence is incomplete.

*We changed the sentence (page 10, 2nd para): “In previous studies, the test-retest reliability for handgrip strength (intraclass correlation coefficient [ICC] = 0.97; 95% confidence interval [CI]: 0.94-0.99), intra-, and interrater reliability were good (ICC = 0.85-0.98) in healthy subjects [18,19].”*

16. You use ‘good’ correlation on p 13, and ‘strong’ correlation, on p 15. Please use one term only.

*We changed (page 13, 3rd para) into ‘strong’.*

Discretionary Revisions

17. Title: using ‘:’ twice and ‘?‘ should be avoided.

*We changed the title.*

18. Language is generally good but needs checking. Examples:

- p4, Results: … was associated to … either ‘with’ or, in context of regression, ‘explained by’.
- p5 last para: … be focusing for…. … beliefs of being not healthy…
- p6 line 2: … attention should focus on …

*We changed the three sentences and the whole manuscript has been edited by a native-English speaker with scientific expertise.*

- line 7: … is not studied before … has not been studied before … should be operationalized… does not seem a focus in this study. Omit sentence?

*We agree. Therefore, we rewrote the introduction section. Thank you for (in our view) improving the quality of our manuscript.*

- ICF Figure: omit DOT (or explain)

*We added the explanation of the DOT to the ICF Figure 1.*
Response to reviewer #2: Martin Descarreaux

We thank you for your valuable comments and thorough review of our manuscript, which in our view has improved the quality of it.

General comments This paper presents a validation of various tests currently used in the functional assessment of workers or future employees (screening tool). Its main purpose is to determine the factors related to the performance in five different physical tests in healthy adults. According to the authors conclusion results in functional tests such as the handling test are mostly explained by physical or personal factors. Moreover, none of the tests were related to psychological factors. Although the presented results are based on a large sample of workers and although I believe that there is a lot of relevant information that could be derived from the generated data, I have doubts regarding how the statistical analyses were conducted (please refer to the Major Compulsory Revisions for details). Until these issues are properly addressed, I cannot fully assess the manuscript. The manuscript should also be revised by an English speaking editor prior to re-submission.

For the preparation of the second draft we were assisted by Prof. Dr. W.P. Krijnen, biostatistician at the Hanze University of Applied Sciences; assistant professor at the University of Groningen.

The manuscript and letter are revised by an English speaking editor.

Major Compulsory Revisions

1. Comment 1. The authors have clearly stated their goal and the purpose of the study which is to “to identify relationships between functional capacity and other ICF factors in a sample of healthy workers, and to determine how much statistical variance in functional capacity can be explained by these factors”. In the methods and results section, however, the authors introduce a new research question by conducting analyses to test “the mean differences in functional capacity between same-gender evaluator worker pairs and mixed-gender evaluator-worker pairs”. The backgrounds, current state of knowledge, purposes as well as a clear hypothesis regarding such analyses need to be addressed in the introduction.

We discussed this in our study group and concluded that evaluator’s gender can indeed be considered as a new research question which can be addressed in a separate study. We removed evaluator’s gender from the factors measured in this study.

2. Comment 2. Evaluations were conducted by inexperienced evaluators (although they received proper training). I don’t believe this invalidates the data but the authors should address this point in the limitation section of the manuscript as results may differ when evaluations are conducted by experienced clinicians or assessors.

We addressed this comment by adding a limitation section (page 19, 2nd para): “Although the evaluators were well instructed in the test protocol, the results of this study may differ from a sample that was evaluated by experienced evaluators.”

3. Comment 3. My major issue with the manuscript is the statistical analyses and their interpretation. In each of their statistical models, the authors included Gender, Weight, Aerobic fitness and Strength. The rational for inclusion is clearly stated in the manuscript. However, I suspect these variables are highly correlated between each other which raise major concerns regarding multicollinearity in these models. Multicollinearity can be defined as a statistical phenomenon in which multiple predictors in a multiple regression model are highly correlated. In such cases the models may not give valid results about any individual predictor, or about which predictors are redundant with respect to others. I don’t feel fully qualified to comment on such issue but I would
highly recommend that the manuscript be screened by a biostatistician prior to publication. At this stage, results obtained from the models where men and women were mixed seem to indicate that men perform differently than women in functional capacity testing. (Gender is a string predictor and men are heavier, have higher aerobic capacity and are usually stronger!)

For the preparation of the second draft, we were assisted by Prof. Dr. W.P. Krijnen, biostatistician. He advised us to use the minimum BIC criterion in order to find a parsimonious and well interpretable regression model. This provides insight in how explanatory variables may explain variance. In particular, whether height and weight as useful as predictors additional to gender. In some cases it did (lifting low and carrying) and in others (standing forward bend, overhead working) gender could be replaced by e.g. weight.

As is indicated in the scatter plots below and in Appendix 2, there is, indeed, a correlation between gender, muscle power, weight, and aerobic fitness. Gender is the strongest significant predictor for the lifting low FC test.

We agree that inter-correlation needed better clarification. Therefore, we added an inter-correlation table (Appendix 2)

We rewrote the statistical analysis section on regression analyzes (page 14, 2nd para).

We added to the result section (page 15, 1st para): “Significant strong inter-correlations were described in appendix 2. Gender is fairly correlated to material handling FC tests and strongly correlated with handgrip strength. Elbow flexion inter-correlated significantly and strong with elbow extension, shoulder abduction, and handgrip strength (Appendix 2).”

We added in the discussion section (page 17, 2nd para): “It may be noted that the models found by minimum BIC are best but do not exclude models explaining little less variance e.g. muscle strength is replaced by another, based on strong inter-correlations (Appendix 2).”
Minor Essential Revisions

- Currently two different titles. Please select one.

  Based on your comments and first reviewer’s comments, we changed the title.

- Manuscript needs to be revised by an English speaking editor

  Our manuscript has been edited by a native-English speaker with scientific expertise.

Abstract section

4. Comment 4. ICF is not defined in the abstract.

  We changed it accordingly (Abstract, page 4).

5. Comment 5. Changes should be done according to the different comments and changes performed throughout the manuscript.

  Thank you for improving our manuscript.

Background section

6. Comment 6. “Lower or higher FC may be related to a range of factors such as gender, muscle strength or personal factors”. Please be more specific and include proper references.

  This sentence was meant as an introduction to the next paragraph. Based on reviewers’ comments we rewrote the background section, whereby the sentence was removed.

7. Comment 7. “Furthermore, more attention should focus on the influence of the test evaluator himself during the test administration as an environmental factor” This sentence is meaningless unless more details regarding the role of the test evaluator are given.

  We removed the sentence.

Methods

8. Comment 8. The authors refer to a detailed description published elsewhere. This sentence stands alone at the end of the sample section of the methods and I cannot identify what it relates to?

  We replaced the sentence from the end of the Method-study sample section to the measures Functional Capacity section (page 9, 1st para): “A detailed description of the FCE test protocol is published elsewhere [15] and can be requested from the corresponding author.”

9. Comment 9. The body function section needs to be clarified. It is currently written in a way that the reader is led to believe that the authors are presenting their results when they are in fact referring to published validity and reliability studies.

  Thank you for attending us on this point. We changed this section (page 10, 2nd para).

Results

10. Comment 10. The Material handling results section needs clarifications. The authors report data comparing subjects who performed “better” in various tests (one standard deviation greater than
the mean). I do not understand the rationale behind these results. The standard deviations and related mean are not presented in the manuscript!

With this sentence we attempted to interpret the model for the readers. Now we agree that this needs further clarification. We rewrote the sentence and verified the clarity of this sentence with clinicians that are not working in the field of research.

The standard deviations were described in table 2.

We have made changes (page 16, 3rd para) and additions (page 17, 1st para) to improve clarity.

Discussion

11. Comment 11. The whole paragraph regarding psychological factors (second complete paragraph page 19) is vague and confusing. The authors refer to various classification and related results which are not described nor discussed in the manuscript.

We changed the wording of the psychological factors, rewrote and rearranged the discussion section in order to be more specific (page 17-20).

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests: 'I declare that I have no competing interests'

References


