**PART A:**

<table>
<thead>
<tr>
<th>Journal Name:</th>
<th>British Journal of Medicine and Medical Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript Number:</td>
<td>MS: 2012/BJMMR/1607</td>
</tr>
<tr>
<td>Title of the Manuscript:</td>
<td>Validity of workers’ self-reports. Evaluation of a question assessing lifetime exposure to occupational physical activity.</td>
</tr>
</tbody>
</table>

*(Note: Title of different sections as proposed below may differ in case of review paper)*
### GENERAL COMMENT:

While this study presents a means of evaluating the validity of self-report measures, the issues identified with the question in stage 1 did not appear to be addressed, and these may have impacted on the results of the other stages. One has to question the point in proceeding to these stages, without ensuring that the question can be understood and completed correctly by the majority of participants.

The need to investigate the validity and reliability of data collection methods for occupational tasks does need to be highlighted and the possible methods for doing so discussed in a broader context to promote the use of valid and reliable measures in studies. This paper may, therefore, be more suited to a paper critically discussing the processes involved in determining the validity of similar questions in ergonomic studies, or the reliability of similar interviews, rather than specifically discussing them in relation to this question.

### SPECIFIC COMMENTS:

There should be critical analysis of the references cited, particularly those which lead to decisions being made about this study.

There are a few grammatical issues throughout which need addressing:

1. ‘..., too.’ Should be ‘... too’ (no comma)
2. References to Tables, Figures etc throughout require capitals
3. Consistency is required in terms of kappa values or kappa-values throughout
4. Consistency is required in terms of ‘sedentary’ or sedentary throughout
5. Occasionally et al is missing from the references when used as part of the sentence
6. There is occasional misuse of capitals (i.e. Cognitive in line 54)

The Danish question included in the Appendix is not clear

### TITLE AND ABSTRACT

The title is appropriate.

Abstract: in the results kappa values are listed in some cases with respectively following them. This does not make sense in these sentences, so it is recommended that the author consider changing these to a range of values, i.e. 0.27-0.29 and 0.60-0.71.

Conclusion: this is a bit ambiguous, as OPA includes sedentary work. Consider changing to ‘...self-reports of exposure to high level OPA are questionable.’ It is also recommended that this conclusion is clearer, stating that this validity assessment only refers to this particular question in the sample investigated.

### INTRODUCTION (IS THE PROBLEM/OBJECTIVE OF THIS

The advantages of questionnaires over interviews is well described, as is the need for valid and reliable measures.
| Study original, important and well defined? | How this study fits within the CAMB project could be expanded in this section, including the aims of the larger project and how this question fits within it, in addition to the pilot testing referred to in 107. The issues identified in the pilot testing and the need for further investigation following this could be expanded in this section.  

A paragraph regarding occupational injuries, and other methods of assessment would be of benefit, in placing this in a broader context (i.e. beyond the CAMB project).  

Given this paper focuses on one question perhaps the question could be included in the background, along with the reference to Lund et al. (2009). Some discussion regarding the use in Lund et al’s (2009) study would provide greater depth, perhaps including issues they faced in using this question.  

The purpose of the three stages and how they link to determining the validity of the self-report question should be discussed, particularly why the validity of the interviews is not examined, and why the reliability of the self-report question is not investigated.  

I would consider clearly stating research questions at the end of this section. While the paper’s intention is outlined in lines 35-36, this lacks depth. This can be overcome by having questions, perhaps linked to the three stages. |
|---|---|
| Are all the references cited relevant, adequate? Are there any other suitable current references authors need to cite? | No problems identified  

No missing references identified |
| Materials & methods  
(Kindly comment on the suitability of the methods. Sufficient details of the methods should be provided to allow peers to evaluate and/or replicate the work) | It appears that the CAMB questionnaire was completed prior to this evaluation study. It would be of benefit to add to this section, or perhaps the background, why the evaluation was not carried out on a smaller scale prior to the full questionnaire being distributed, and what indicated that this needed evaluating following the questionnaire (i.e. were there unexpected results, missing data etc). I appreciate that it was piloted but there was no information reported regarding whether changes were made following this pilot testing, or why there still appeared to be issues for participants completing it despite this pilot testing.  

Table 1 needs CEBM questionnaire added to step 2 methods  

Section 2.2.1 consider stating that sampling to redundancy was used prior to describing the included participants  

Section 2.2.2 the QAS checklist may be better presented as dot points or in a table  

Section 2.3 in line 136 either describe the interview guide here, or state that it was developed for this study based on stage 1 |
Section 2.3.1 please justify the sample size

Was the judgement of the OPA discussed with the participant at all at this stage, or an interview summary agreed with to improve the validity of the interview process? Given the issue of validity of these interviews is stated in the background, it needs to be addressed if being used as a reference measure.

Please describe the CAMB sample, and how this sample was obtained from within it. For instance were the 15 from each group randomly selected from all of those who were eligible.

Section 2.3.2 Please define 'exposed' and 'non-exposed'. It does read as any exposure (i.e. any duration/time) would be classed as exposed, however, I wonder whether there was a degree of exposure (i.e. at least 2 years).

It may be worth including the Landis and Koch interpretations of kappa so that the reader does not need to find these. Could you also justify why you chose these interpretations?

Please reference the interpretation of the Bland-Altman plots (lines 174-175)

Section 2.4.1.2 Please justify the sample size

Section 2.4.2.2. Please explain why this Bland-Altman plot has a y of 0 with no error bars
Results & discussion
(Kindly comment on: 1. Are the data well controlled and robust? 2. Authors should provide relevant references during discussion. 3. Discussion and conclusions should be based on actual facts and figures. Biased claims should be pointed out. 4. Are statistical analyses must for this paper? If yes, have sufficient and appropriate statistical analyses been carried out?)

Section 3.1.1 participant demographics may be better placed in section 2.2.1 with the other demographic info regarding this group.

Section 3.1.2 the primary issue from the results appeared to be the categories, and the interpretation of the questions, yet one sentence refers to this. Furthermore this sentence only comments on the sedentary jobs being easily categorised. I would consider restructuring this section to focus more on this. Perhaps:
- Issues with categories, including suggestions to overcoming these issues, or discussion regarding similar questions from the literature
- Issues with recall. This information is sufficient
- Limitations - what you have with regards to sample size and semi-structured interviews, however some references to the literature would strengthen your argument regarding these not being an issue
- Benefits of interview

The last sentence is ambiguous. Do you mean that it should be used in conjunction with valid and reliable measures? If so, expand on why? This doesn’t really fit in this section as this section is about the process of completing the questionnaire.

Section 3.2.1. I would remove ‘few participants had a high OPA index’ from line 251 as it doesn’t fit in terms of the flow of the paragraph, and is essentially stated in line 253.

The term passive is ambiguous. Either state that it is sedentary (if that’s what you mean) or clearly define this term.

Section 3.2.2. is there any literature about non-sedentary tasks? If so, please add it, or state that there isn’t any. The statement ‘our results are in line with others...’ reads as though this is the case of all intensity levels, but isn’t backed up with references from the literature. If only sedentary tasks have similar results in the literature this has essentially already been stated (line 257).

Were any unintended results obtained in this stage? If not, why do you think the question was better completed in this stage, in comparison with stage 1. If there were unintended results, what was the proportion? How were these dealt with (i.e. were they excluded) Where they more common in a particular sub-group within the sample?

Consider rewriting the paragraph lines 265-270. In the statement citing White et al in what way are they superior? Why do you think they are more valid? Why did you not attempt to test the validity of the interview? The last sentence of this paragraph does not make sense.

Paragraph lines 271-282. Did you collect data regarding socio-demographics, age, gender, education etc? If so this needs to be reported. Otherwise, please make it clear that these are assumptions being made and not necessarily the case for this sample. Please also back these assumptions with
references to the literature (if possible). Was data collected regarding conditions/fitness? If so please report this.

Limitations of this stage need to be discussed. This may include the lack of demographic data collected (if this is the case), clearly stating that there may be issues with the inter-method validation especially where the validity of the interviews is not considered, any unintended results. You may also talk about the issues outlined in stage 1, as these may relate to the validity, i.e. if participants found it difficult to categories more active tasks then that may account for the validity issues noted at higher intensities in this section.

Section 3.3.2 I’m not sure why validity is being discussed in this section?

I question the statement regarding the reliability of self-reported. The reliability of self-reported exposure is likely to be affected by the questionnaire, particularly where the participants have difficulties in understanding and completing the question. For instance, if someone has difficulty deciding which category their job fits into they could decide to go with one category on one occasion, but another on the second occasion.

Nevertheless the reference to Stock et al. 2005 should be in the background, along with a critical analysis of the study, to justify why the reliability of the self report measures is not being investigated in this study.

Add to the discussion the issues, possible reasons for, and possible solutions for the reliability at higher OPA indices.

Please describe the pitfalls you refer to in expert judgements.
### SDI Review Form 1.5

<table>
<thead>
<tr>
<th>Are there any ethical or competing interest issues?</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conclusion</strong> (Is the conclusion supported by the data, discussed inside the manuscript? Conclusions should not be biased and should be based on the data, presented inside the manuscript only)</td>
<td>This included reference to the Job Exposure Matrix which has not been mentioned in the study previously. It is not clear why qualitative methods used in validating survey questions will also be useful in occupational medicine. While this may be true, this is not supported by the results. The only qualitative analysis was in stage 1, where only the processes involved in completing the questionnaire are obtained through the interview, hence this statement is not supported. The issues with the validity of the self-ratings for non-sedentary tasks should be added to the conclusion. This needs to be stated with respect to the cohort (i.e. those working for 20+ years).</td>
</tr>
</tbody>
</table>

**References**
The two Torgen et al. (1999) studies are currently referenced identically in-text.

---

**Note:** Anonymous Reviewer: Reviewer requested not to reveal his/her identity.