Author's response to reviews

Title: Measuring walking within and outside the neighborhood in Chinese elders: reliability and validity

Authors:

- Ester Cerin (ecerin@hku.hk)
- Anthony Barnett (abarnett@hku.hk)
- Cindy HP Sit (sithp@hku.hk)
- Martin Cheung (martin_cheung@dh.gov.hk)
- Janet Lee (h0460544@hku.hk)
- Daniel SY Ho (syho@hku.hk)
- Wai-man Chan (w_m_chan@dh.gov.hk)

Version: 2 Date: 17 August 2011

Author's response to reviews:

We thank the Reviewers and Associate Editor for their comments on our manuscript. We also thank the Associate Editor for her interest in our manuscript, encouraging comments, and for giving us the opportunity to revise our manuscript in light of the Reviewers' comments.

We have now addressed all concerns raised by Drs. Sharpe and Zhu. These are detailed below.

Response to Reviewers' and Associate Editor's comments:

Reviewer 1: Dr. Patricia Sharpe

This study among Chinese elders in Hong Kong has the multiple aims of adapting the walking section of a Neighborhood Walking questionnaire, assessing reliability and validity, and examining the potential impact of neighborhood SES and walking purpose on reliability and validity. Overall the paper is well-written and the methods are rigorous.

Our reply:

We thank Dr. Sharpe for her positive appraisal of our manuscript.

Minor essential revisions

1. An area of confusion to the reader appears on page 5 in the last sentence of the section:

All participants were asked if they would consent to wearing an accelerometer for a week, keeping a log of walks and being re-assessed on the questionnaire two weeks after the first assessment. From the pool of consenting participants (73% of the original sample), three per street block (n=96) were randomly selected for this component of the study.

It is not clear to what research activities "this component" is referring to. Do you
mean the “main” study as opposed to the pilot study?
Our reply:
Yes, this is correct. We were referring to the main study. We have now omitted the phrase “this component of” from the sentence, as it was creating confusion rather than clarifying what we meant to say. Now the relevant section reads:
“For the main study, participants were recruited using a two-stage sampling strategy, whereby street blocks with at least 25 EHCs residing members were randomly selected without replacement in each of the four catchment areas. ... All participants were asked if they would consent to wearing an accelerometer for a week, keeping a log of walks and being re-assessed on the questionnaire two weeks after the first assessment. From the pool of consenting participants (73% of the original 484 participants), three per street block (n=96) were randomly selected to take part in this study. No significant differences were found between consenting and non-consenting participants …”

2. On page 6: “On day 8 of the study, the accelerometers and log were collected, the data were checked for validity and the NWQ-CS was interviewer-administered to those with at least five valid days of data, including a weekend day”.
By stating “the data were checked for validity”, do you mean that the accelerometer wear data were downloaded and checked to see if the wear time was adequate for inclusion in the subsequent analyses? Better precision in what you are saying here would aid the reader.
Our reply:
Yes, this is correct. We have now clarified that:
“On day 8 of the study, the accelerometers and log were collected. Accelerometer data were downloaded and checked to see if the daily wear-time was adequate for inclusion in subsequent analyses. Logs of walks were also checked for completeness (e.g., provision of all requested information). The NWQ-CS was interviewer-administered to those with at least five valid days of data, including a weekend day. Nine participants who failed to meet the validity criteria were asked to extend the data collection for another week and their interview was rescheduled accordingly.”

3. On page 7: Please put a published citation or website to the WHO guidelines for translation.
Our reply:
We have now included the relevant WHO webpage.
“The Chinese version of the walking section of the NPAQ (NWQ-CS) was initially translated from English to Chinese and back-translated from Chinese to English following the World Health Organization guidelines (www.who.int/substance_abuse/research_tools/translation/en).”

4. Discussion: It is confusing and distracting in the Discussion to compare findings of this study to those of another study by Cerin et al., that is not yet
published, which employed the IPAQ. As the reader does not have access to that study’s results, they should not be discussed in the manuscript.

Our reply:

The manuscript in question (reporting on the IPAQ) has in the meanwhile been assessed as needing “minor revisions” to be acceptable for publication in the Journal of Aging and Physical Activity (JAPA). We have resubmitted a revised version addressing the reviewers’ and associate editor’s comments in July 2011. We thus believe that there is a high chance that this manuscript will be soon available to the readership. We think that the comparison between the findings related to the two instruments (IPAQ vs. NWQ-CS) is quite interesting and informative. We leave it to the Editors to decide whether this information should be omitted. An alternative course of action would be to wait for a final decision on the JAPA submission before this manuscript is finally accepted for publication (obviously, assuming that it is suitable for publication in BMC Public Health).

5. Additionally, the Discussion states as a limitation the following: “Limitations of this study include the use of accelerometer cut-points developed for a younger population (adults) in absence of established cut-points for older adults. However, the Measures section states the following: “Previously published cut-points, that were employed in earlier studies with older adults [23-24], were used to classify activity counts into light (100-1951 counts/min), moderate (1952-5724 counts/min) and vigorous (>5724 counts/min) [25-26].” Which statement is correct? Were the cut-points previously used with older adults or not? Is there evidence of their validity among older adults?

Our reply:

These cut-points were used in studies with older adults. However, they were developed using samples of adults. Therefore, their validity among older adults is questionable. This has been identified as a research area that definitely needs further work. To avoid confusion, we have now reworded the statement in the Methods section as follows:

“In absence of cut-points validated in older adults, previously published cut-points originally developed for adults were used to classify activity counts into light (100-1951 counts/min), moderate (1952-5724 counts/min) and vigorous (>5724 counts/min) [24-25]. Despite their unknown validity, these cut-points have also been employed in earlier studies with older adults [23, 26].”

6. There are a few omitted words and other minor errors throughout.

Our reply:

We have now carefully checked the manuscript for typos and corrected them.

Discretionary revisions:

7. On page 8: It is unclear what the use of % agreement for reliability of categorical variables adds to this study beyond the use of kappa statistic. Percent agreement has the limitation of not taking into account the agreement that may occur by chance. The kappa statistic corrects for this (Szklo and Nieto,

Our reply:
Percent agreement was used jointly with kappa statistic because it is not negatively affected by low levels of variability in the data. If there is no variability in the participants' responses (e.g., everyone gives an affirmative answer on a checklist item on two different assessments), kappa statistic cannot be computed. In this case, the agreement between assessments will be 100%. Low variability in responses may also yield low kappa values (which is due to how it is computed), while percent agreement may be very high (>90%). We have now added a sentence to the data analytic plan that provides an explanation for our decision to compute both percent agreement and kappa statistic.

“The variability of categorical variables was assessed using Kappa statistic and percent agreement. The latter was computed because Kappa statistics can be low and suggest poor reliability when there is little variability in the responses [29].”

Reviewer 2: Dr. Weimo Zhu
Using a good sampling control, a relatively large sample and a valid objective physical activity measure, the authors made a welcome attempt to cross-validate NWQ-CS, a neighborhood physical activity questionnaire, with a focus on walking.

Our reply:
We thank Dr. Zhu for his appreciation of these aspects of our study.

1. From the measurement perspective, however, there are several major limitations of this study:
Criterion measure was poorly selected. The authors used ActiGraph and 7-day walking log as the criteria measures to validate NWQ-CS. Since the interest of the study is to determine if NWQ-CS can “differentiate whether walking occurs within or outside of the neighborhood, ActiGraph cannot be served as the criterion measure since it provides no information on the location. The location devices, like GPS, should be employed.

Our reply:
Unfortunately, the current GPS technology does produce reliable and valid location data in Hong Kong. As we had noted in the limitations section of our original submission and has been noted by Dr. Ainsworth (Associate Editor):

“It would have been optimal to assess walking settings using Global Positioning System (GPS) monitors [40]. However, the built environment of Hong Kong with its urban canyons, steep hills, and underground walkways poses significant challenges to GPS data collection because the GPS signal is often blocked and the available satellite signals are insufficient to estimate the positioning information [41]."
Therefore, we applied the best possible available criterion measures (diary of walks and accelerometry) for field studies suitable to the Hong Kong urban environment. We would also like to stress that accelerometer data were only compared against total walking rather than walking within and outside the neighborhood since, as Dr. Zhu noted, accelerometers cannot capture walking location (within vs. outside the neighborhood).

2. The results did not support the validity conclusion. Log-of-walks employed can be served as a criterion measure, but it generated a much higher total weekly min of walking for transport than NWQ-CS, and some correlations between these measures were only at the moderate levels (see Table 4). However, the authors concluded that “This study provides support for the validity and reliability …,” which is an incorrect statement.

Our reply:
We believe that our conclusions on the validity and reliability of the NWQ-CS were a bit more subtle, as Dr. Ainsworth highlighted in her comment. In the original submission, we stated that:

“… the NWQ-CS was a valid measure of walking for recreation and frequency of walking for transport. With regards to duration of walking for transport, the NWQ-CS may systematically underestimate the actual level of walking in samples that engage in high levels of this type of walking. Nonetheless, it appears to be able to reliably pick up individual differences since the correlations between the NWQ-CS and log corresponding measures were moderate (r=0.41 and 0.56) and close to the validity cut-off values considered to be acceptable (r=0.50 for comparisons with diaries) [32].”

In the conclusion of the original submission, we also stated:

“However, caution is needed in interpreting estimates of transport-related walking duration, as the examined sample tended to underestimate the amount of walking they did. Future studies will need to explore ways to minimize the observed bias. This might include the provision of clearer instructions to respondents that highlight the need to report regular and occasional trips to destinations.”

3. “Score reliability” rather than “instrument reliability” was studied. While the interest of the study is on the reliability of NWQ-CS, the authors employed a typical “score reliability” data collection design: After NWQ-CS was administered first time, the measure was administered to the participants two weeks later. Since between-time differences are the variation of participants' walking behavior, the reported reliability is indeed mainly the “stability” of participants' walking behaviors. Since this design mixed up both instrument reliability and behavior stability, the reliability observed is really the “score reliability.” For a behavior survey, the degree of recall error is likely the major “reliability” interest. Therefore, “retest” should be done within 24 hours.
Our reply:

We do not agree with Dr. Zhu that what we assessed was in the main the stability of participants’ walking behavior. The NWQ-CS measures ‘usual’ not ‘last week’ walking, which by definition should be stable within a 2-to-3-week period. This would have been different if the instruction had been to report the frequency and amount of walking undertaken in the last week or last day. We also see a major problem with test-rested protocols that involve 24-hour intervals. Namely, they are unlikely to be able to differentiate the degree of recall of a specific behavior we are interested in from the degree of recall of the answer given to a questionnaire item. It would be interesting to see how these two sources of reliability could be disentangled.

4. Some of the data analysis and reporting look fuzzy. As an example, the authors stated that: “differences in validity estimates between areas were evaluated by including appropriate interaction terms in the regression models”. There is no description why this analysis should be the validity evidence of NWQ-CS. Then, when reporting the results, there is only one sentence on this analysis: “No differences in validity estimates were observed between types of neighborhood.” There is NO any statistical result to back up the statement, which is not acceptable for a scientific paper.

Our reply:

The assessment of differences in reliability and validity estimates between areas were meant to provide information on whether the reliability and validity of self-reports of walking behavior differ by neighborhood walkability and neighborhood SES. This was explained in the original manuscript as follows:

“No studies have examined whether the test-retest reliability and validity of the NPAQ differ across types of neighborhoods. The NPAQ is an instrument particularly relevant to studies on environment-walking associations where context plays a major role and respondents are sampled from multiple neighborhoods with specific characteristics. Residents of walkable neighborhoods may show higher levels of reliability because they more regularly and frequently engage in walking [12, 19]. Residents of areas with lower socio-economic status (SES) are likely to be less educated and have difficulties in understanding questionnaire items, which may result in a greater error of measurement when compared to their higher SES counterparts [20]. Such between-neighborhood differences in metric characteristics may yield spurious between-neighborhood differences in environment-walking associations. Thus, it is important to know whether neighborhood characteristics moderate the reliability and validity of the NPAQ.”

With regards to Dr. Zhu’s observation that we did not report the results of the analyses related to the moderating effects of area walkability and SES on the reliability and validity of the NWQ-CS, we wish to note that this would have added another table (on non-significant) results to our manuscript and made the manuscript more cluttered. We can add this information in a subsequent revision,
if the Associate Editor deems it is necessary. To address Dr. Zhu’s comment, we have now expanded the sentence reporting no areas moderating effects as follows:

“No differences in validity estimates were observed between types of neighborhood (p-values of area by criterion measure interaction terms > .680”).

5. In addition, the manuscript was poorly written and organized. As an example, the description of “Measures” should be presented before the data collection procedures. It is not clear why title and vol. of a journal article in the reference are in bold. The current manuscript makes little contribution to the literature. Quality of written English: Not suitable for publication unless extensively edited

Our reply:
At least three native English-speakers (Drs. Sharpe - Reviewer, Ainsworth – Associate Editor and Barnett – Co-author) maintain that this manuscript is well-written. I would tend to trust their judgment. Nonetheless, we have thoroughly re-reviewed the manuscript and corrected it for typos.

The fact that the description of “Measures” was not presented before the data collection is due to one of these measures being adapted and developed in the present study. Therefore, we thought that it would make more sense to talk about the various aspects of the study first (pilot and main validation study) before talking about the adapted instrument (NWQ-CS). We understand that this does not follow the APA style, which Dr. Zhu may be more accustomed to. However, the BMC Public Health journal does not follow the APA style.

As to the title and volume of journal articles in the reference being in bold, this was done because it is a formatting requirement of BMC Public Health.

As to the comment that this manuscript makes little contribution to the field, we wish to stress that:

The aims of this study were to:

(1) Adapt the walking section of the NPAQ for Chinese-speaking elders (which did not exist prior to this work)

(2) Examine its reliability (no prior data were available on this issue);

(3) Assess its validity evidence against accelerometry-based estimates of PA and daily logs of walking (this is the first study to publish validity data on the NPAQ in any country / population group)

(4) Determine the moderating effects of neighborhood-level SES and transport-related walkability on reliability and validity estimates (never done before in any population, although it is an important issue to better understand eventual differential association of environmental factors with PA outcomes).