Reviewer’s report

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

Version: 1 Date: 9 August 2011

Reviewer: Weimo Zhu

Reviewer’s report:

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. From the measurement perspective, however, there are several major limitations of this study:

1. 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.

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.

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.

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.
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.

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

**Quality of written English:** Not suitable for publication unless extensively edited

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

**Declaration of competing interests:**

I declare that I have no competing interests.