Journal of Physical Activity and Health, 2012, 9(Suppl 1), S29-S36

A Checklist for Evaluating the Methodological Quality of Validation Studies on Self-Report Instruments for Physical Activity and Sedentary Behavior

Maria Hagströmer, Barbara E. Ainsworth, Lydia Kwak, and Heather R. Bowles

Context: The quality of methodological papers assessing physical activity instruments depends upon the rigor of a study's design. ***Objectives:*** We present a checklist to assess key criteria for instrument validation studies. ***Process:*** A Medline/PubMed search was performed to identify guidelines for evaluating the methodological quality of instrument validation studies. Based upon the literature, a pilot version of a checklist was developed consisting of 21 items with 3 subscales: 1) quality of the reported data (9 items: assess whether the reported information is sufficient to make an unbiased assessment of the findings); 2) external validity of the results (3 items: assess the extent to which the findings are generalizable); 3) internal validity of the study (9 items: assess the rigor of the study design). The checklist was tested for interrater reliability and feasibility with 6 raters. ***Findings:*** Raters viewed the checklist as helpful for reviewing studies. They suggested minor wording changes for 8 items to clarify intent. One item was divided into 2 items for a total of 22 items. ***Discussion:*** Checklists may be useful to assess the quality of studies designed to validate physical activity instruments. Future research should test checklist internal consistency, test-retest reliability, and criterion validity.

Keywords: external validity, internal validity, reporting, questionnaire

Assessing physical activity is important for prevalence estimates; establishing trends in physical activity levels and sedentary behaviors over time; establishing the link between different aspects of physical activity, sedentary behavior, and health; and for evaluation of behavior change as part of a physical activity or lifestyle intervention program.[1] Valid and reliable methods are needed as poor assessment of physical activity may lead to inadequate conclusions.[1,2] Multiple physical activity and sedentary behavior self-report instruments have been developed and are found in the literature.[2,3]

The ever-increasing volume of information about the measurement properties of the self-report instruments has potential to guide the evolution of the methods; however, it is difficult to discern which aspects of questionnaires assess physical activity with acceptable precision without rigorous evaluation of the research methods applied in validation studies. In 2010, van Poppel et al performed a review of measurement properties of available self-report physical activity instruments.[4] They found the overall quality of the methodological papers was poor and in need of improvement. Only a few questionnaires had sufficient construct validity and reliability to identify physical activity and sedentary behaviors as intended. Responsiveness and sensitivity to change was only studied in 2 of the 85 questionnaires evaluated. The authors concluded that more attention should be paid to the methodology of studies assessing measurement properties of physical activity instruments and the quality of reporting the validity and reliability studies. Potential reasons for their conclusion might be a) the general lack of guidance on how to develop and test a physical activity or sedentary behavior instrument, b) how to choose a self-report instrument from the many available, and c) how to assess the quality of existing validation studies.

One way to guide physical activity researchers with the identification of the most appropriate and valid instrument for their research purposes is to provide a checklist that assesses the quality of existing validation studies. This checklist should focus on 3 important issues, namely the quality of the reported data and the external and internal validity. This will allow the reader to make an unbiased assessment of the findings of the study. By including items on external validity, a judgment can be made whether the findings from the study can be generalized to the population from which the study subjects were derived. Factors influencing internal validity include selection of appropriate study sample and measurement error.

Hagströmer is with the Dept of Neurobiology, Care Sciences, and Society, Karolinska Institute, Huddinge, Sweden. Ainsworth is with the Program in Exercise and Wellness, College of Nursing and Health Innovation, Arizona State University, Phoenix, AZ. Kwak is with the Dept of Biosciences, Karolinska Institute, Huddinge, Sweden. Bowles is with the Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.

In the last decade, guidelines for reporting and assessing the quality of different study designs have been developed. Such guidelines include the CONSORT (Consolidated Standards of Reporting Trials) Statement with recommendations on how to report randomized controlled trials (RCT),[5,6] the STROBE Statement (Strengthening the Reporting of Observational Studies in Epidemiology) with recommendations on how to report observational studies,[7] the Downs and Black checklist for assessment of quality of randomized and nonrandomized studies[8] and the Delphi list for quality assessment of randomized clinical trials.[9] These checklists are, however, not specifically aimed at assessing the quality of physical activity or sedentary behavior instrument validation studies. A checklist that judges the appropriateness of physical activity validation methods is the one developed by Rennie and Wareham in 1998.[10] Their checklist evaluates the strengths and limitations of physical activity validation studies and focuses on dimensions of physical activity, comparison methods, choice of study populations, mode of administration, and statistical considerations. However, it does not cover all aspects of needed to assess study quality, such as internal and external validity. No checklists have been developed since then to assess the quality of physical activity validation methods.

Objectives

The purpose of this study is to propose a checklist, named the Hagströmer-Bowles Physical Activity/Sedentary Behavior Questionnaire Checklist (abbreviated as, HBQC), which aims to assess key methodological quality criteria for physical activity and sedentary behaviors self-report validation studies. The paper will outline the development of the checklist, including the justification and explanation of the included items, the results of the pretest of the checklist and the final refined version of the checklist.

Process

Development of the HBQC Checklist

The first step in the development of the checklist was a Medline/PubMed search to identify published guidelines or standards for evaluating the methodological quality of validation studies on self-report measures of physical activity and sedentary behavior. The Rennie and Wareham checklist was the only set of principles we found that was created explicitly for physical activity instruments.[10] Their checklist was used to further refine our search of the electronic database and to categorize our findings. Moreover, we examined published reviews and compilations of measures to identify additional criteria to include in our checklist. A major finding of our search is the overall agreement in the literature on 1 key criterion for physical activity validation studies, namely that the physical activity construct purported to be measured by the instrument should be clearly defined. The importance of a clear definition of the physical activity constructs used has also been highlighted by others. For example, in an examination of the methodological quality of physical activity and coronary heart disease studies, Powell et al wrote "The concept and operational definition of physical activity is vital to the value of these studies."[1] For further information about physical activity construct see the article in this supplement by Pettee-Gabriel, Morrow, and Woolsey.

The HBQC was developed based upon the Downs and Black (1998) checklist for measuring study quality[8] and on additional methodological criteria specific to questionnaire design and physical activity and sedentary behaviors assessment identified in the literature.[1,11–23] The Downs and Black checklist has 27 items that assess reporting, external validity, internal validity and power in relation to randomized and nonrandomized studies. We modified 20 of Down and Black's items to focus on physical activity for use in the HBQC. The remaining items were developed from information gleaned from the literature search. The pilot version of the HBQC consisted of 21 items and contained 3 distinct subscales that address the following items:

1. *Quality of the Reported Data*: This subscale has 9 items which assess whether the information provided in the paper is sufficient to allow a reader to make unbiased assessment of the findings of the study.
2. *External Validity of the Results*: This subscale has 3 items which address the extent to which the findings from the study can be generalized to the population from which the study subjects were derived.
3. *Internal validity of the Study*: This subscale has 9 items which address methodological bias, bias related to study sample and to the power to prevent type 1 and type 2 errors.

Justification of Items Included in the HBQC

Each of the 21 items on the pilot version of the checklist are explained and justified below.

Subscale 1: Quality of the Reported Data

Item 1. For each type of research including methodological studies a clear and well defined hypothesis, aim, or objective is crucial.[15]

Item 2. The key constructs of physical activity and/or sedentary behaviors used in the instrument should be clearly defined in the introduction or method section. Further, it is imperative that the constructs defined also are the ones evaluated.[10] This is of great importance as some papers state that the questionnaire is valid for assessing intensity, frequency, and duration of physical activity as well as sedentary time even though the questionnaire has only been tested for agreement between

total physical activity and energy expenditure and not sedentary time.

Items 3 and 4. A description of the participants and principal confounders such as sex, age, body mass index (BMI), education, and race/ethnicity should be provided.[11,20,21] This information will inform the reader whether the study was performed on convenience samples, which often consists of young, highly educated subjects, or if the instrument is valid for other diverse groups.

Item 5. For transparency in a validation study, it is crucial to describe how and on which groups the original instrument was developed and tested. If modifications to the instrument occur, such as language translations, cultural adaptation, and changes in the order of the questions, then a justification is needed to explain how and why the changes occurred. If modifications have been made to the questionnaire this might influence the validity of the modified questionnaire.[13,18,22]

Item 6. The reference measure, defined as the physical activity/sedentary behavior measures used to compare the self-report instrument, needs to be clearly described, both how and what it measures as well as how data reduction was done. Different ways to handle the data, so called data reduction, will lead to different results (ie, it has to be transparent how the data are handled as the mode of questionnaire administration, such as self-administered and interview format, can have effects on data quality).[13]

Item 7. To assess the quality of the results it is important to describe the characteristics of participants with missing, incomplete or invalid data. This will inform the reader whether a specific group has difficulties in completing the questionnaire. It is possible that more complex and longer a questionnaires increase the risks of having missing, incomplete or invalid data.

Item 8. Methodological studies on physical activity and sedentary behavior questionnaires should include subjects with a wide range of physical activity levels and sedentary behaviors,[10] and they should report the variability of the data. For validation studies on physical activity measures it is important to include subjects with a wide range of physical activity levels to test if the scale/measure is valid for both low, moderately, and highly active individuals.

Item 9. The limits of agreement (Bland-Atman plots) should be reported when comparing ratio scales. Nevill and Atkinson (1997) conducted measurement agreement in 13 studies (23 examples) using variables recorded on a ratio scale in sports medicine and sports science.[19] The results provided strong evidence that heteroscedastic errors are the norm with ratio scales. The authors concluded that if the correlation between the absolute measurement differences and the mean of the instruments is positive and greater than the equivalent correlation using log transformed measurements, the ratio limits of agreement will provide a more precise comparison of the instruments than other measures.

Subscale 2: External Validity of the Data

Item 10. The study must identify the source population for the sample and describe how the sample was selected. The study population selected for a validation study should reflect the population to whom the questionnaire will be applied (ie, both men and women, different age groups, disease groups, national representative, or different minority groups).[10,11,18,20] The use of a convenience sample, such as university staff introduces a bias as it is neither representative of the general population nor of the group that the instrument is intended to be used on.

Item 11. The proportion of participants excluded (or nonresponders) after recruitment should be stated. Information whether the enrolled group of participants is representative should be provided. Exclusion criteria should be examined as it may exclude the intended group for which the instrument was developed.

Item 12. The survey mode, such as using self-report, telephone or face-to-face interview, should be reported. In addition, the instrument needs to be validated using the same mode of administration as intended for the instrument when developed and as it used in research settings. A review comparing different measures of physical activity showed that the same instrument delivered by self-report and interview gave very different results.[21]

Subscale 3: Internal Validity of the Study

Item 13. Physical activity is a complex behavior. One technique to change the behavior is to raise awareness of one's physical activity level. By filling in or responding to a questionnaire, awareness is raised and the behavior can be altered and thereby influence the results.[13,24,25] The same is true if the burden for the individual to fill in the questionnaire and do other tests is high as they might report more poorly on the instrument, skip items, or drop-out of the study. It is therefore important to make efforts to reduce the burden for the participants.[13,20]

Item 14. Information should be provided whether an attempt was made to blind the research staff, as there is a risk for bias between persons collecting the data. For example, interviewers can introduce bias by encouraging respondents to give specific answers on a survey instrument or technicians performing a fitness test may encourage some participants to work harder or stop a test early depending on respondent characteristics.

Item 15. A problem in validation studies of physical activity instruments is to select an appropriate reference measure.[10] Often it is not possible to find a gold standard that measures the same behaviors without error, so another instrument needs to be selected. The reference method chosen should have a higher precision than the

method tested and a high correlation with a method closest to a gold standard.[26]

Item 16. The time frame of the questionnaire validated needs to be considered when choosing a comparison method, as they should assess physical activity in the same time frame.[10] If the time frame is 1 year or one's usual behavior, repeated measures over a year should be performed. If the time frame is last 7 days the same period of time should be measured using the reference method (ie, temporally matched).

Item 17. Another issue that seems obvious for research but sometimes not followed is compliance to the measurement protocol. Study procedures should follow the study protocol.

Item 18. Test-retest reliability and responsiveness to behavior change is crucial if the instrument is going to be used to monitor physical activity or sedentary behaviors over time or to evaluate an intervention.[14,17]

Item 19. The approach of analysis is dependent on the type and outcome measured, such as on an ordinal scale or continues scale, and whether it is reduced to a continuous ordered categorical variable, or discreet categorical variable. In most cases physical activity data are positively skewed, which means that other methods than parametric analysis needs to be employed or data transformation using parametric or nonparametric techniques.[10,12] In a 2006 review of statistical methods used for the validation of questionnaires, the authors concluded that the correlations were a common statistical approach and this was concerning since correlations can yield misleading conclusions in validation studies.[23] Further, Bland-Altman methods are recommended to improve epidemiological data quality by providing a graphical representation of agreement between a criterion measure and the questionnaire.[10,12] It should be used as a complement of statistical tests of agreement using for example Intra Class Correlation for continues variables and kappa statistics for ordinal variables. Such estimates would help to increase the precision of estimates when physical activity is being used as predictor to examine disease outcome relationships in epidemiology studies.

Item 20. Any analyses that had not been planned at the outset of the study should be clearly indicated. This item is related to data dredging which can be defined as the inappropriate use of data mining to uncover misleading relationships in data or data analyzed to seek for any possible relationships.

Item 21. For scientific and ethical reasons, the sample size for a study needs to be planned, with a balance between clinical, epidemiological, and statistical considerations.[5–7,10] Note that the sample size should be calculated in accordance with the study aim [ie, if the aim is to detect agreement for principal confounder subgroups (eg, men and women separately) the sample size calculation needs to take that in to account].

Scoring the HBQC

To keep the checklist straightforward and to be seen as an indicator of quality we decided to use a simple scoring system. Subscale items are scored 0 (no or unable to determine) or 1 (yes), except for 1 item regarding sample size in the internal validity subscale that is scored 0 (no), 1 (partially), and 2 (yes). Each of the 3 subscales on the HBQC is scored with the subscales summed to provide a total checklist score. Providing subscale scores helps to identify strengths and weaknesses of each methodological concern addressed in the checklist. The total maximum score for the pilot version was 22.

Pretesting the HBQC

Before assessing the interrater reliability and feasibility of using the HBQC, a senior physical activity researcher (BA) and a medical statistician were asked to comment on the checklist after which some modifications were made. Modifications included minor wording and item content changes. In addition, 2 physical activity measurement experts attending the "Measurement of Physical Activity: Closing the Gaps in Self-Report Methods" (referred to as "Self-Report") meeting in July 2010 provided comments on the checklist.

Assessment of the Interrater Reliability and Feasibility of the Pilot Version of the HBQC

Using different raters, the checklist was tested for interrater reliability and feasibility. Doctoral students, Post-Doctoral and research fellows attending the "Self-Report" meeting in July 2010 were asked to participate in an evaluation of the pilot version of the HBQC. Six persons (4 PhD students and 2 Post-Doctoral fellows) scored 2 or 3 different physical activity methodological studies and 2 PhD students scored 2 different sedentary behavior methodological studies. The papers scored were selected from validation studies published in the last 4 years. The selection was done by one of the authors to select papers assessing both continues and ordinal outcome variables. For the assessment of methodological papers on physical activity we used the paper by Egeland et al, titled "Concurrent validity of the International Physical Activity Questionnaire (IPAQ) in the Cree Community, Canada,"[27] which was scored by 5 different raters. The paper by Besson et al, titled "Validity of the historical adulthood physical activity questionnaire,"[28] was scored by 4 raters. Two persons scored the paper by Macfarlane et al, titled "Reliability and validity of the Chinese version of the IPAQ."[29] For the sedentary behavior review, the papers by Marshall et al, titled "Measuring total and domain–specific sitting: a study of reliability and validity,"[30] and by Otten et al, titled "Relationships between self-report and objective measure of television-viewing time in adults,"[31] were scored by 2 different raters.

Findings

Findings of Interrater Reliability

Table 1 describes the subscales scores and total score per paper evaluated. In general, the raters found the HBQC helpful and easy to use. They gave some comments on how to improve the clarity and understanding of the checklist. The raters were quite consistent in their scoring of the subscales reporting and internal validity. For external validity, item 13 (regarding minimizing awareness and burden for the participants), item 14 (regarding blinding research staff), and item 21 (regarding sample size power) seemed most difficult to interpret and score.

Refining the HBQC

Given the results from the reliability test, comments from the experts and the different raters the HBQC was revised slightly and is presented in the Appendix. Minor wording changes were made to 8 of the items to clarify the intent of the checklist item and 1 item addressing 2 different concepts was divided into 2 items. While the initial checklist had 21 items, the revised checklist now has 22 items (Section A = 10 items; Section B = 2 items; Section C = 10 items) and the total score is 23.

Next Step

Given the evolving state of the technology it may be difficult or inappropriate to develop a static tool or system. Our results suggest there is agreement on key criteria. The next step is to identify and assess a clear, simple set of consensus criteria that the general public can understand and use.

Conclusions

In this paper we have described the development of a checklist for assessing key criteria for physical activity and sedentary behavior validation studies. The checklist has 22 items and 3 subscales that address how study results are reported, external validity, and internal validity. These issues are important for researchers planning method testing studies as well as for those who plan to apply physical activity measurement in new studies or interpret results of already published studies. Future research should test the internal consistency, test-retest reliability, criterion validity, and respondent burden of the checklist.

Acknowledgments

The authors would like to thank Stephen Herrmann from Arizona State University, Miguel Andrés Calabró from Iowa State University, Brigid Lynch from Alberta Health Services, Calgary, Matthew Buman from Stanford University, Elin Johansson from Karolinska Institutet, Sweden, Ing-Mari Dohrn from Karolinska Institutet, Sweden and Browyn Clark from University of Queensland, Australia for volunteering in the inter-rater reliability testing and also giving useful feedback on the feasibility of the checklist.

References

1. Powell KE, Thompson PD, Caspersen CJ, Kendrick JS. Physical activity and the incidence of coronary heart disease. *Annu Rev Public Health*. 1987;8:253–287.
2. Mahar K, Ainsworth BE. Measurement of physical activity. *Res Quarterly Exerc Sport*. 2000;71:1–158 (Special Issue).
3. Kriska A, Caspersen CJE. A collection of physical activity questionnaires for health-related research. *Med Sci Sports Exerc*. 1997;29(6 Suppl):S5-S205.
4. van Poppel MN, Chinapaw MJ, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. *Sports Med*. 2010;40(7):565–600.
5. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
6. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.

Table 1 Median and Range of the Rating by Each Paper Scored Using the Pilot Version of the Hagströmer-Bowles Physical Activity/Sedentary Behavior Questionnaire Checklist (HBQC)

	Besson et al* (N = 5)***	Egeland et al* (N = 4)***	Macfarlane et al* (N = 2)***	Marshall et al** (N = 2)***	Otten et al** (N = 2)***
Subscale A: Reporting (max 9 points)	7 (6–8)	5 (4–8)	6.5 (6–7)	8.5 (8–9)	6.5 (6–7)
Subscale B: Internal Validity (max 3 points)	1 (1–2)	1.5 (1–2)	1 (1–1)	1.5 (1–2)	0.5 (0–1)
Subscale C: External Validity (max 10 points)	6 (5–7)	5 (3–6)	7 (6–8)	6 (5–7)	4 (3–5)
Total score (max 22 points)	14 (13–17)	11.5 (10–14)	14.5 (13–16)	16 (14–18)	11 (11–11)

* Physical Activity Methodological Studies.

** Sedentary Behavior Methodological Studies.

*** Number of raters.

7. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61(4):344–349.
8. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52(6):377–384.
9. Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol*. 1998;51(12):1235–1241.
10. Rennie KL, Wareham NJ. The validation of physical activity instruments for measuring energy expenditure: problems and pitfalls. *Public Health Nutr*. 1998;1(4):265–271.
11. Ainsworth BE, Sternfeld B, Slattery ML, Daguise V, Zahm SH. Physical activity and breast cancer: evaluation of physical activity assessment methods. *Cancer*. 1998;83(3, Suppl):611–620.
12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.
13. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Public Health (Oxf)*. 2005;27(3):281–291.
14. Davidson M, de Morton N. A systematic review of the Human Activity Profile. *Clin Rehabil*. 2007;21(2):151–162.
15. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830–837.
16. Lagerros YT, Lagiou P. Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. *Eur J Epidemiol*. 2007;22(6):353–362.
17. Le Grande MR, Elliott PC, Worcester MU, Murphy BM, Goble AJ. An evaluation of self-report physical activity instruments used in studies involving cardiac patients. *J Cardiopulm Rehabil Prev*. 2008;28(6):358–369.
18. Martinez SM, Ainsworth BE, Elder JP. A review of physical activity measures used among US Latinos: guidelines for developing culturally appropriate measures. *Ann Behav Med*. 2008;36(2):195–207.
19. Nevill AM, Atkinson G. Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. *Br J Sports Med*. 1997;31(4):314–318.
20. Pols MA, Peeters PH, Kemper HC, Grobbee DE. Methodological aspects of physical activity assessment in epidemiological studies. *Eur J Epidemiol*. 1998;14(1):63–70.
21. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act*. 2008;5:56.
22. Rattray J, Jones MC. Essential elements of questionnaire design and development. *J Clin Nurs*. 2007;16(2):234–243.
23. Schmidt ME, Steindorf K. Statistical methods for the validation of questionnaires—discrepancy between theory and practice. *Methods Inf Med*. 2006;45(4):409–413.
24. van Sluijs EM, Griffin SJ, van Poppel MN. A cross-sectional study of awareness of physical activity: associations with personal, behavioral and psychosocial factors. *Int J Behav Nutr Phys Act*. 2007;4:53.
25. Weinstein LB, Abrams RA, Ayers CS. Increasing awareness of sugar ingestion among children. *Pediatr Nurs*. 1988;14(4):277–279.
26. Wong MY, Day NE, Wareham NJ. Measurement error in epidemiology: the design of validation studies II: bivariate situation. *Stat Med*. 1999;18(21):2831–2845.
27. Egeland GM, Denomme D, Lejeune P, Pereg D. Concurrent validity of the International Physical Activity Questionnaire (IPAQ) in an liyiyiu Aschii (Cree) community. *Can J Public Health*. 2008;99(4):307–310.
28. Besson H, Harwood CA, Ekelund U, et al. Validation of the historical adulthood physical activity questionnaire (HAPAQ) against objective measurements of physical activity. *Int J Behav Nutr Phys Act*. 2010;7:54.
29. Macfarlane DJ, Lee CC, Ho EY, Chan KL, Chan DT. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). *J Sci Med Sport*. 2007;10(1):45–51.
30. Marshall AL, Miller YD, Burton NW, Brown WJ. Measuring total and domain-specific sitting: a study of reliability and validity. *Med Sci Sports Exerc*. 2010;42(6):1094–1102.
31. Otten JJ, Littenberg B, Harvey-Berino JR. Relationship between self-report and an objective measure of television-viewing time in adults. *Obesity (Silver Spring)*. 2010;18(6):1273–1275.

Appendix

Revised Evaluation Template for Assessing Quality of Physical Activity Validation Studies: The Hagströmer-Bowles Physical Activity/Sedentary Behavior Questionnaire Checklist (abbreviated as, HBQC)

Adapted from Downs and Black (1998) checklist for measuring study quality with additional methodological criteria specific to questionnaire design and physical activity assessment identified in the literature.

Subscale A: Reporting (10 Possible Points)

1. Is the hypothesis/aim/objective of the study clearly described?

Yes	1
No	0

2. Are the operational definitions of main physical activity constructs to be validated clearly described in the Introduction or Methods section?

If the main constructs (for example, frequency, duration, intensity, volume, type, domain for physical activity or type of sedentary behavior etc.) are first mentioned in the Results section, the question should be answered no.

Yes	1
No	0

3. Are the characteristics of the participants included in the study clearly described?

Inclusion and/or exclusion criteria should be listed.

Yes	1
No	0

4. Are the distributions of principal confounders clearly described?

A list of principal confounders is provided. At a minimum, distributions of sex, age, and education should be presented. Other confounders, such as BMI or race/ethnicity, may be important depending on the population.

Yes	1
No	0

5. For studies validating an existing self-report measure has the original source of the measure been cited? For studies validating a modified version an existing self-report measure, has the original source of the measure been cited, and have the modifications been clearly described? If the study is validating a new measure the question should be answered yes.

Modifications include cultural adaptation, language translation, change in question order or wording, and change in scoring protocol from the original source. If the study is validating an existing measure developed for a population culturally or linguistically different from the study population without reporting whether the measure was adapted/translated the question should be answered as unable to determine.

Yes	1
No	0
Unable to determine	0

6. Are the methods of administration for the self report measure and the reference measure clearly described?

Yes	1
No	0

7. Are the methods of data reduction for the self report measure and the reference measure clearly described?

Yes	1
No	0

8. Have the characteristics of participants with missing, incomplete, and/or invalid data been described?

This should be answered yes where there were no participant exclusions based on missing or poor data or where the number of exclusions were so small that findings would be unaffected by inclusion. This should be answered no where a study does not describe or report the number of participants excluded based on missing or poor data.

Yes	1
No	0

9. Does the study provide information about the variability in the data for the main physical activity constructs? Simple descriptive statistics (eg, means, standard deviations, medians, value ranges, frequencies) should be reported for both the self report measure and the reference measure so the reader can check the major analyses and conclusions. This question does not cover statistical tests which are considered below.

Yes	1
No	0

10. Have limits of agreement and/or confidence intervals been reported for the main analyses?

Yes	1
No	0

Subscale B: External Validity (3 Possible Points)

11. Were the individuals asked to participate in the study representative of the entire population from which they were recruited?

The study must identify the source population for the sample and describe how the sample was selected. Samples would be representative if they comprised the entire source population, an unselected group of consecutive patients in a clinical setting, or a random sample. Random sampling is only feasible where a list of all members of the relevant population exists. Where a study does not report the proportion of the source population from which the samples are derived, the question should be answered as unable to determine.

Yes	1
No	0
Unable to determine	0

12. Were those participants who were enrolled in the study representative of the entire population from which they were recruited?

The proportion of those excluded after recruitment should be stated. Evidence that the enrolled participants are representative would include demonstrating that the distribution of the principal confounding factors was the same between the study participants and recruited individuals who were not enrolled (excluded and/or nonresponders).

Yes	1
No	0
Unable to determine	0

13. Was the mode of administration representative of the procedures applied for similar study designs?

Yes	1
No	0
Unable to determine	0

Subscale C: Internal Validity—Bias (10 Possible Points)

14. Was an attempt made to minimize altered physical activity behavior by the participant in the use of the validity measures?

For studies where self-report physical activity was assessed by recall and/or participant burden for collecting the reference measure was low this should be answered yes.

Yes	1
No	0
Unable to determine	0

15. Was an attempt made to blind research staff to the activity levels or characteristics of the participants to avoid biasing the results?

For studies where the research staff had no way of knowing the activity levels or characteristics of the participants, or where the self-report measure was administered by self-completion without aid, prompting, or verification by research staff this should be answered yes.

Yes	1
No	0
Unable to determine	0

16. Does the reference measure assess the physical activity construct(s) of interest with greater accuracy than the self-report measure, and are errors in the reference measure uncorrelated with errors in the self-report measure?

For studies where the validity of the reference measure was clearly described, the question should be answered yes. For studies which refer to other work that demonstrates the reference measure is more valid compared with self-report the question should be answered yes. For studies that used a reference measure that is not a direct measure of physical activity (eg, body composition, respiratory function, cholesterol) the question should be answered no.

Yes	1
No	0
Unable to determine	0

17. Did the self-report measure and the reference measure assess physical activity in the same time frame?

If the self-report measure assessed physical activity engaged in over a long period of time during the past, such as past month or past year, the reference measure should have been administered repeatedly over this interval to account for variation in activity.

Yes	1
No	0
Unable to determine	0

18. Was compliance with the measurement protocol acceptable?

Where there was a high level of noncompliance (>20% of sample) or difference in the level of compliance by principal confounder subgroups the question should be answered no. For studies where the effect of any misclassification was likely to bias measures of agreement to the null, the question should be answered yes.

Yes	1
No	0
Unable to determine	0

19. Was reproducibility of the main physical activity construct(s) reported for the self-report measure?

For test-retest reliability was the time interval between self-report measure administrations appropriate?

Yes	1
No	0
Unable to determine	0

20. Were the statistical tests used appropriate to assess agreement for the main physical activity constructs between the self-report measure and the reference measure?

The statistical techniques used must be appropriate to the data. If the self-report and reference measures assess physical activity constructs in the same units (eg, minutes per week) then the analysis should provide an indication of over- or underestimation by self-report (eg, Bland-Altman method). If the measures assess different constructs (eg, minutes per week versus energy expenditure) then correlation coefficients should be provided, or responses to the self-report should be divided into quantiles and reference measure means should be compared. For categorical data sensitivity and specificity should be calculated.

Yes	1
No	0
Unable to determine	0

21. Were all analyses planned at the outset of the study? If data dredging was used, it was identified and made clear. Any analyses that had not been planned at the outset of the study should be clearly indicated. If no retrospective unplanned subgroup analyses were reported, then answer yes.

Yes	1
No	0
Unable to determine	0

22. Did the study appear to have sufficient sample size to assess agreement?

If sample sizes were calculated to detect a level of correlation or agreement for principal confounder subgroups (eg, men and women separately) then the question should be answered as yes. If sample sizes were calculated to detect a level of correlation or agreement for the total sample but not subgroups the question should be answered as partially.

Yes	2
Partially	1
No	0
Unable to determine	0