Title: Reliability and validity of subjective measures of aerobic intensity in adults with spinal cord injury: a systematic review

ABSTRACT

Objective: To systematically synthesize and appraise research regarding test-retest reliability or criterion validity of subjective measures for assessing aerobic exercise intensity in adults with spinal cord injury (SCI).

Data Sources: Electronic databases (Pubmed, PsychINFO, SPORTDiscus, EMBASE and CINAHL) were searched from inception to 1-1-2016.

Study Selection: Studies involving at least 50% of participants with SCI who performed an aerobic exercise test that included measurement of subjective and objective intensity based on test-retest reliability or criterion validity protocols.

Data Extraction: Characteristics were extracted on study design, measures, participants, protocols, and results. Each study was evaluated for risk of bias based on strength of the study design and a quality checklist score (COnsensus-based Standards for the selection of health Measurement Instruments [COSMIN]).

Data Synthesis: The seven eligible studies (one for reliability, six for validity) evaluated overall, peripheral and/or central ratings of perceived exertion on a 6-20 scale (RPE 6-20). No eligible studies were identified for other subjective intensity measures. The evidence for reliability and validity were synthesized separately for each measure, and assessed using Grading of Recommendations Assessment, Development, and Evaluation (GRADE). Overall, very low GRADE confidence ratings were established for reliability and validity evidence generalizable to the entire population with SCI and various upper-body and lower-body modalities. There was low confidence for the evidence showing that overall RPE 6-20 has acceptable validity for adults with SCI and high fitness levels performing moderate to vigorous-intensity upper-body aerobic exercise.

Conclusions: Health care professionals and scientists need to be aware of the very low to low confidence in the evidence, which currently prohibits a strong clinical recommendation for the use of subjective measures for assessing aerobic exercise intensity in adults with SCI. However, a tentative, conditional recommendation regarding overall RPE 6-20 seems applicable depending on participants' fitness level as well as the exercise intensity and modality used.

MeSH Key Words: paraplegia; quadriplegia; spinal cord injuries; exercise; sports

LIST OF ABBREVIATIONS

COSMIN = COnsensus-based Standards for the selection of health Measurement INstruments

GRADE = Grading of Recommendations Assessment, Development, and Evaluation

CR10 = ratings of perceived exertion on a category-ratio 0-10 scale

HR = heart rate

ICC = intraclass correlation

$\dot{V}O_2$ = oxygen uptake

PA = physical activity

Physical Activity Recall Assessment for People with Spinal Cord Injury (PARA-SCI)

PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RPE = ratings of perceived exertion

RPE 6-20 = ratings of perceived exertion on a 6-20 scale

SCI = spinal cord injury

INTRODUCTION

World-wide statistics show that each year between 250,000 and 500,000 people incur a spinal cord injury (SCI) [1]. As a result of profound physical, environmental and psychological barriers to physical activity (PA) participation [2,3], adults with SCI are more physically inactive and deconditioned compared to the general population and other disability groups [4,5,6]. These factors contribute to the increased risk in the SCI population of chronic conditions such as cardiometabolic disease [7,8,9,10].

As a fundamental step toward promoting physical activity (PA) among adults with SCI, the first evidence-based, SCI-specific PA guidelines were developed in 2011 [11]. These guidelines were underpinned by a systematic review and appraisal of evidence regarding the effects of exercise training on fitness of adults with SCI [12]. That review showed that 20 min of moderate to vigorous aerobic exercise, performed twice per week at an intensity of 60-65% peak oxygen uptake ($\dot{V}O_2$) or 60-80% peak heart rate (HR) is required for adults with SCI to gain important fitness benefits. Such fitness benefits have been positively associated to health, participation and quality of life of adults with SCI [13,14,15]. However, $\dot{V}O_2$ and HR measures of exercise intensity cannot be used by many adults with SCI. The cost of $\dot{V}O_2$ equipment is prohibitive for most rehabilitation centers and exercise environments in the community [11], while sympathetic decentralization renders HR to be an unsuitable method for assessing aerobic intensity in those with lesion levels at or above the fifth thoracic vertebra [16,17,18].

Subjective measures of aerobic intensity are considered reliable and valid alternatives to $\dot{V}O_2$ and HR for assessing exercise intensity within the able-bodied population [19,20]. Such measures are based on the psychological integration of cardiorespiratory, musculoskeletal and metabolic signals of exertion, into ratings of perceived exertion (RPE) using, for example, a 6-20 scale (RPE 6-20) or a 0-10 category-ratio scale (CR10) [21].

However, the able-bodied evidence cannot be generalized to the SCI population. The interpretation of signals of exertion might be altered by impaired afferent feedback from the exercising muscles, a decentralized sympathetic nervous system, and/or peripheral fatigue of the small active muscle mass during upper-body aerobic exercise [19,21,22,23,24].

Notwithstanding, both RPE 6-20 and CR10 have been used to assess exercise intensity in aerobic exercise interventions for adults with SCI [25,26,27,28,29,30,31,32,33,34,35,36,37]. Furthermore, recent data suggests that using differentiated RPE could improve the assessment of upper-body aerobic exercise intensity compared to the traditional overall RPE [38,39]; differentiated RPE involves separately rating peripheral RPE (signals from the exercising limbs) and central RPE (cardiorespiratory signals), instead of using overall RPE (integrated rating of the peripheral and central signals). Another subjective intensity measure suggested to adults with SCI is a PA intensity classification chart [11], part of a reliable and valid SCI-specific PA questionnaire (Physical Activity Recall Assessment for People with Spinal Cord Injury [PARA-SCI]) [40].

However, it is not yet clear whether adults with SCI can use these different subjective measures in a reliable and valid fashion to assess intensity during various forms of aerobic exercise. If so, this would provide the evidence base for adults with SCI to self-regulate exercise intensity without $\dot{V}O_2$ or HR measures. These questions warrant a systematic review on the fundamental measurement properties of test-retest reliability and criterion validity [41]. Protocols to test these measurement properties for subjective intensity measures have previously been developed (Table 1) [20,42]. Accordingly, the purpose of this systematic review was to synthesize and appraise research regarding test-retest reliability or criterion validity of subjective measures for assessing aerobic exercise intensity in adults with SCI.

METHODS

The conduct and reporting of this review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) [43]. The review protocol was not registered.

Bibliographic databases and keywords

The following electronic bibliographic databases were searched for studies published from inception up to 1-1-2016: Pubmed, PsychINFO (EBSCOhost), SPORTDiscus (EBSCOhost), EMBASE (OVID), CINAHL (OVID). Databases were searched by combining keywords representing SCI with keywords representing subjective exercise intensity (Supplement 1). Language was restricted to English [44].

Study eligibility criteria

Studies were included if:

- at least 50% of the participants were adults (≥16 years) with traumatic or non-traumatic SCI, excluding those with spina bifida or multiple sclerosis;
- participants performed an aerobic cyclic exercise test (e.g. arm cranking, wheelchair propulsion, bodyweight-supported ambulation) of at least 3 min in which a subjective intensity measure was used simultaneously with measurement of $\dot{V}O_2$ or HR [45,46] and
- a reliability and/or validity protocol was used in accordance with Tables 1 and 2 [20,42].

Peer-reviewed studies with single-case and group designs were included. Studies or individual data were excluded if solely based on HR in participants with lesions levels at or above the fifth thoracic vertebra, in whom a decentralized sympathetic nervous system renders HR to be potentially unsuitable for assessing exercise intensity [16,17,18].

Eligibility screening

Two reviewers (XXXXX and XX) conducted eligibility screening independently, while not being blinded to authors or journals. The citations identified through the database searches were combined and duplicates were removed (Figure 1). The reviewers then scanned titles and abstracts, excluding citations that clearly did not met eligibility criteria. Following this, the full-texts of the remaining citations were reviewed; non-eligible citations were excluded while recording reasons for exclusion. Finally, the reviewers scanned reference lists of included studies for potentially eligible citations not identified through the database searches. Differences were identified at all stages between the reviewers, who then reached a final decision by together re-reviewing the title, abstract and/or full text against the eligibity criteria.

Data extraction

One reviewer (XXXXX) extracted data from the included studies, verified by a second reviewer (XX). Data extraction (Table 3) included pre-allocated fields on:

- the subjective measure evaluated (e.g. overall RPE 6-20, peripheral RPE 6-20, CR10);
- participant characteristics (i.e., demographics, lesion characteristics, fitness levels, and PA levels);
- study protocol (i.e., test protocol, exercise modality, exercise intensity, familiarization with the subjective measure, and if/how the subjective measure was prompted during exercise); and
- results (i.e., individual or group data of subjective intensity and $\dot{V}O_2$ or HR as well as statistics on reliability or validity).

Following this, the benchmarks shown in Table 2 were used to assess if the results of each study indicated acceptable, unacceptable or inconclusive test-retest reliability or criterion validity for the evaluated subjective measure. The benchmarks were based on PA questionnaire studies [47], and the assumption that >10% variation in $\dot{V}O_2$ or HR is unacceptable for a subjective intensity measure to be considered reliable or valid.

Risk of bias of each study

One reviewer (XX) assessed risk of bias of each study, verified by a second reviewer (XXXXX). Quality of each study was assessed using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist, which has been developed through a transparent and rigorous process [48]. The checklist was considered applicable given that subjective aerobic intensity measures bear many resemblances to health measurement instruments. The COSMIN checklist includes a section with 14 items on test-retest reliability and a section with seven items on criterion validity. The items on statistics required modification in accordance with Table 2 (see Supplement 2 for the modified items). The lowest rating of any of the items within a section defined the overall score for each included study, which could be “Excellent”, “Good”, “Fair”, or “Poor”. After verification by XXXXXX, two items required further discussion between the reviewers: appropriateness of the time interval (item #8 for reliability) and whether there were “minor” or “major” flaws in the study designs (item #10 for reliability, item #5 for validity). The COSMIN criteria were re-evaluated to reach a final decision.

A level of evidence was then designated for each study based on the quality score and, for validity studies, strength of the study design. Level 1 reliability studies were studies of Excellent or Good quality, while Level 2 reliability studies were studies of Fair or Poor quality. Level 1 and 2 validity studies were based on an estimation-production design (Level 1: Excellent or Good quality; Level 2: Fair or Poor quality). The single-test relationships design was considered a weaker design than the estimation-production procedure for assessing the criterion validity of assessing aerobic intensity using a subjective measure [49].

Accordingly, validity studies using a single-test relationship design were designated as Level 3 (Excellent or Good quality) or Level 4 (Fair or Poor quality).

Synthesis and appraisal of evidence

For each subjective intensity measure, an evidence summary was drafted for studies that showed acceptable, unacceptable, or inconclusive reliability/validity. Each evidence summary included descriptive data on quality scores, participant characteristics, exercise modality, exercise intensity, familiarization with the subjective measure, and if/how the subjective measure was prompted during exercise (Table 4).

These summaries were then used to assess the evidence for each measure using Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) [50,51]. The GRADE method prescribes assessing the evidence for risk of bias, inconsistency, imprecision, indirectness, and publication bias [50,51]. If one or more of those issues appeared, the GRADE confidence ratings was downgraded from "High" to "Moderate", "Low" or "Very low" [50,51]. Benchmarks for these criteria were developed for this review (Supplement 3). A "Very serious" risk of bias was defined by a lack of Level 1 or 2 studies, and "Serious" risk of bias by the presence of only one Level 1 or 2 study. Inconsistency was defined by less than two third of studies showing acceptable reliability/validity (Table 2). Imprecision was assessed based on (i) the absence of adequately powered studies (for 80% power to detect an $ICC \geq 0.70$ or $r \geq 0.80$, using a one-tailed test with $\alpha = 0.05$, $N \geq 11$ or $N \geq 8$ are the minimal sample sizes, respectively [52]), and/or (ii) more than half of the studies providing inconclusive results due to large interindividual differences (Table 2). Indirectness was defined by the evidence not including study groups representative of the SCI population as well as various exercise modalities and intensities (Supplement 3). Publication bias was considered absent, based on scanning reference lists and searching trial registers (Supplement

3). A higher GRADE confidence rating [50,51] was considered if excellent reliability/validity (e.g. ICC ≥0.80) was found across the majority of studies (Supplement 3).

Finally, for each intensity measure, a Conclusion based on the GRADE assessment was formulated regarding the confidence in the evidence. These statements reflected the generalizability of the evidence towards the entire population with SCI (e.g. acute and chronic SCI, physically active and inactive), towards various exercise intensities (light, moderate and vigorous), and towards the various upper and lower-body exercise modalities applicable for adults with SCI [12].

RESULTS

From 647 unique citations, seven studies were found eligible; one for test-retest reliability and six for criterion validity (Figure 1). These seven studies evaluated overall, peripheral and/or central RPE 6-20 (Table 3). Eligible studies on other subjective measures of aerobic intensity were not identified.

Studies regarding test-retest reliability

In the one reliability study, overall RPE 6-20 was assessed in 102 participants with acute SCI [53]. The study included men and women with varying PA levels and lesion characteristics, in whom predominantly peak $\dot{V}O_2$ <1.00 $L \cdot min^{-1}$ was found. No details were provided on procedures for familiarizing participants with the use of RPE. It was reported that RPE was prompted visually and verbally during exercise. The group performed two maximal arm crank or wheelchair ergometry tests separated by eight weeks. Under those conditions, the reported ICC of 0.47 indicated that reliability of overall RPE 6-20 was unacceptable. However, it was not clear whether findings were confounded by changes in the participants occurring between the test and retest (e.g. neural recovery, improved upper-body skills). The study therefore received only a "Fair" quality rating (Supplement 2).

Synthesis and appraisal of evidence regarding test-retest reliability

A Very low GRADE confidence rating in the evidence was established through the GRADE assessment for three reasons. Firstly, there was Serious risk of bias, given there was only one Level 2 study. Secondly, there was a lack of directness for the population (absence of adults with lumbar lesions or chronic SCI) and protocols used (no evidence for light and moderate exercise intensities and modalities other than upper-body exercise). Finally, it had to be assumed that the evidence lacks precision, as no ICC confidence intervals or limits of agreement were presented. Accordingly, the Conclusion was formulated as: "There is very low confidence in the evidence evaluating the reliability of overall RPE 6-20 for adults with acute SCI performing maximal-intensity upper-body exercise, and therefore also very low confidence for evidence regarding other SCI populations, exercise intensities and modalities."

Studies regarding criterion validity

In the six eligible validity studies, overall RPE 6-20 was used in five studies [22,54,55,56,57] peripheral RPE 6-20 in three studies [22,55,58], and central RPE 6-20 in two studies (Table 3) [22,55]. Two studies [22,54] used an estimation-production design consisting of a 20-min $\dot{V}O2$-regulated trial that was reproduced based on RPE. Five studies [22,55,56,57,58] used a single-test relationship design to establish the correlation between $\dot{V}O_2$ and RPE during a maximal or submaximal test. One study used an estimation-production as well as a single-test relationship design [22]. Data were only reported or eligible for $\dot{V}O_2$ (Table 3), except for one study that included data on Pearson's *r* between RPE 6-20 and HR [55].

Four out of the six studies included adults with chronic SCI and high fitness levels who performed sports at an elite or recreational level (37 out of 50 total participants)

[22,54,55,58]. In the other two studies, PA and/or fitness levels were not reported [56,57]. Across the six studies, adults with various lesion and completeness levels were included, but not women or adults with acute SCI.

Five studies employed various upper-body modalities (wheelchair ergometry [22,55], arm crank ergometry [58], hand cycle ergometry [54], and recumbent stepping [57]), while the sixth study utilized electrically-stimulated ambulation [56]. Moderate and/or vigorous intensities were assessed in the two studies that used an estimation-production design [22,54], while light, moderate and vigorous intensities were evaluated in the studies employing a single-test relationship design. In four out of the six studies, RPE was prompted visually and verbally during exercise after receiving detailed verbal instructions on how to use the RPE scale [22,54,55,58]. Details on these methods were not provided in the other two reports [56,57].

All studies received a Fair or Poor quality rating owing to inappropriate use of statistics (e.g. no Fisher transformation when averaging Pearson's r), minor flaws in the design of the study (e.g. potential selection bias), and/or inadequately powered samples (Table 4 and Supplement 2). Assessment of the checklist items for each study can be found in Supplement 2.

Synthesis and appraisal of evidence regarding criterion validity

Overall RPE 6-20: The limits of agreement of the two Level 2 studies [22,54] indicated that most, but not all participants were able to use overall RPE 6-20 to reproduce 50 and/or 70% peak $\dot{V}O_2$ with a relative difference <10% (Table 3). The Level 4 studies [22,55,57,58] suggested that overall RPE 6-20 was strongly correlated to $\dot{V}O_2$ in all but one participant performing upper-body exercise, while lower correlations were found among participants performing ambulation [56] (Tables 3 and 4). This lack of consistency and precision, along with the absence of study groups representative of the entire SCI population, led to a Very low GRADE confidence rating in the evidence (Table 4). However, there was no indirectness for adults with chronic SCI and high fitness levels performing upper-body exercise at a moderate to vigorous intensity (50-70% peak $\dot{V}O_2$ and RPE 12-16), after receiving verbal instructions about overall RPE 6-20, and while being prompted visually and verbally with the RPE 6-20 scale during exercise. Accordingly, for that evidence, a conclusion reflecting slightly higher (but still low) confidence was formulated (Table 4).

Peripheral RPE 6-20: Although the three studies [22,55,58] indicated acceptable validity for peripheral RPE 6-20, all were Level 4 studies. This lack of higher-quality studies, along with a lack of directness for the SCI population and various exercise modalities, led to a Very low GRADE confidence rating in the evidence (Table 4). The lack of higher-quality studies prohibited a conclusion reflecting higher confidence in the evidence for a subgroup under specific conditions, in contrast to overall RPE 6-20 (Table 4).

Central RPE 6-20: The two studies [22,55] indicated acceptable validity for this measure, but both were Level 4 studies. The GRADE assessment revealed similar limitations in the evidence as those for peripheral RPE 6-20, again leading to a conclusion reflecting Very low confidence in the evidence (Table 4).

<u>DISCUSSION</u>

This systematic review is the first to synthesize and appraise evidence regarding the test-retest reliability and criterion validity of subjective intensity measures for assessing aerobic exercise intensity in adults with SCI. Through our rigorous and transparent approach in accordance with standards for developing clinical guidelines [50,59], the review provides health care professionals and scientists with the information required to make evidence-based decisions [60] for assessing aerobic intensity in adults with SCI. This approach also allowed identification of the most imminent research matters, as discussed below.

Evidence regarding test-retest reliability

The only eligible reliability study was a lower-quality study evaluating overall RPE 6-20 in adults with acute SCI performing maximal-intensity upper-body exercise. This therefore resulted in there being very little confidence in the evidence regarding test-retest reliability. This is in stark contract with able-bodied research, in which several studies have shown acceptable test-retest reliability for the use of RPE in assessing exercise intensity [19]. However, these studies did indicate that between-trial reliability of RPE to assess intensity increases from the second to the third trial, compared to the first to second trial [19]. This implies that participants need familiarization using an exercise test to reliably self-assess exercise intensity using RPE, and suggests practice improves the reliable use of RPE [19]. Only two trials were conducted in the reliability study included in this review, which could explain the low ICC in that study, of 0.47 [53]. Another confounding factor may have been the eight-week period between test and retest. In this period, neurological recovery of afferent feedback [3] or changes in upper-body skills [61] may have influenced assessment of RPE of the participants [19,22], who had only recently incurred SCI.

This very limited evidence base highlights issues to be addressed in future research. First, high-quality reliability studies are required that include participants with chronic SCI, various exercise intensities, and various exercise modalities. Second, the influence of familiarization and practice on RPE estimates needs to be investigated, i.e., to determine how much practice is needed to yield reliable RPE. Third, there is no evidence of measures other than overall RPE 6-20 specifically assessing the test-retest reliability of an aerobic exercise bout in accordance with appropriate designs (Table 1). Although reliability studies were identified for other subjective intensity measures [40,62], these did not use a study design eligible for evaluating subjective intensity during aerobic exercise (Table 1). For example, acceptable test-retest reliability has been found in an adequately-powered study regarding the intensity classification chart of the PARA-SCI [40]. However, because the PARA-SCI is a self-report measure of overall PA and leisure-time PA [40], the test-retest protocol for the intensity classification chart involved recalling the intensity of activities, rather than reporting the intensities of aerobic exercise bouts as they occurred. Another study indicated acceptable reliability for a subjective measure to assess wheelchair racing intensity, but it was ineligible for this review as >50% of participants had disabilities other than SCI [62]. Finally, quality could be improved by applying standard reporting criteria based on Table 1 and the COSMIN checklist (Supplement 2); examples are improved reporting of statistical methods, how missing data were handled, and provision of individual data to allow additional analyses by others, if necessary.

Evidence regarding criterion validity

The review identified promising evidence indicating that overall RPE 6-20 may have acceptable validity for adults with chronic SCI and high fitness levels performing moderate to vigorous-intensity upper-body aerobic exercise. However, there can still be no more than low confidence in that evidence due to the lack of precise, consistent results. Although there was consistent evidence for peripheral and central RPE 6-20, it was based on lower-quality studies, leading to very low confidence in that evidence.

Significant gaps in knowledge remain for validly assessing aerobic exercise intensity using subjective measures in adults with SCI, as the quality and size of the current SCI evidence lags far behind that for the general population [19,20]. These gaps can be addressed in several ways. First, adequately-powered, high-quality studies using estimation-production designs are required that not only include participants with high fitness levels but also physically inactive or deconditioned adults with SCI who are found in the far majority of the SCI population [4,5,6]. Presumably, physical inactivity or deconditioning imply less experience with exercise and the sensations connected to subjective intensity, which may reduce the valid use of RPE [19,20]. Thus the ability to assess exercise intensity using RPE with acceptable validity could be different based on PA level.

Second, high-quality studies are required to assess if and how reliability and validity of subjective measures of intensity are influenced by lack of afferent feedback from the exercising limbs during clinically popular exercise modalities such as functionally electrical stimulated cycling and ambulation exercise [63]. It also remains to be investigated whether reliability and validity differ among upper-body exercise modalities such as arm cranking and wheelchair propulsion, for example due to differences in mechanical efficiency [39].

Third, the validity evidence for aerobic exercise is currently limited to RPE 6-20. Validity studies regarding other measures have been conducted [40,64], but were not based on an eligible study design for aerobic exercise (Table 1). For example, acceptable validity has been found in an adequately-powered study regarding the intensity classification chart of the PARA-SCI [40], but this finding was based on *recalling* one day of overall PA during which $\dot{V}O_2$ data had been collected, as opposed to reporting the subjective intensity during the activity. Another example was a study regarding the validity of the Talk Test for assessing exercise intensity in adults with SCI [64]. This study was considered ineligible for this review given that its protocol for the estimation trial (maximal exercise test) was not matched with the production trial (20-min exercise bout). Furthermore, Borg's CR10 has been used in various SCI exercise interventions [25,26,27,28,29,30,31,32,33,34,35,36,37], but there is no reliability or validity data to support the use of this measure in adults with SCI performing aerobic exercise. Most of these interventions showed positive effects of exercise on fitness and health when prescribing a range of CR10 aerobic intensities (3 to 7). However, there was little to no information provided on how the CR10 was employed, what the actual objective and subjective intensities were during the exercise sessions, and whether these responses changed over the training period. The current intervention research can therefore not be used to recommend a specific subjective intensity to improve fitness and health.

Fourth, the evidence base could be supported by availability of data of individual participants. This may for example allow calculation of appropriate statistics (Table 1), or recalculation of otherwise ineligible data, which for instance may have allowed the inclusion of an adequately-powered validity study that used absolute $\dot{V}O_2$, instead of the required % peak $\dot{V}O_2$ [65]. Another example is providing data of CR10, along with $\dot{V}O_2$ and HR, of individuals performing a maximal exercise test as part of an intervention. In a future analysis, these data could be used to assess validity in accordance with the single-test relationship design (Table 1).

Finally, improved reporting in accordance with Table 2 and the COSMIN checklist shown in Supplement 2 would strengthen the evidence base. Quality of the evidence could also improve if all future studies reported if and how participants were familiarized with a subjective intensity measure, and how the measure was prompted during exercise, which may be another factor influencing the validity of subjective intensity measures [21].

Study limitations

It is possible that there is evidence from non-English literature that was not captured by this review, but this seems unlikely based on previous reviews [44]. Furthermore, we considered contacting authors for additional data, for example to improve data quality of some studies through conducting appropriate statistical analyses. However, this was not considered resource-effective; the other quality issues for these studies would still have led to the same COSMIN quality scores and GRADE assessments.

Recommendations for practice

Based on the GRADE framework for moving from evidence to recommendations [66], health care professionals and scientists need to be aware that a strong clinical recommendation for the use of subjective measures of aerobic intensity is prohibited considering the lack of moderate or high-quality evidence. However, a tentative, conditional recommendation seems appropriate for the emerging evidence base for overall RPE 6-20, since it is supported by the positive judgement regarding the other domains of the GRADE framework, i.e. estimates of values and preferences, resource use, and the balance between desirable and undesirable outcomes (see Supplement 4 for an overview). There is data showing the high value placed on subjective measures of exercise intensity by adults with SCI and health care professionals [11]. In addition, resources required to implement subjective intensity measures are much lower than costly alternatives such as $\dot{V}O_2$ monitoring. The balance between potential desirable and undesirable outcomes is also positive. A subjective measure of aerobic intensity could support important fitness improvements, while the only undesirable outcome is underestimation of actual intensity leading to more vigorous exercise. This may be an acceptable risk assuming the participant has no contraindications to vigorous exercise based on consultation by a health care professional [11].

Accordingly, the following conditional recommendation may be provided to health care professionals and scientists making evidence-based decisions for assessing aerobic intensity in adults with SCI: “Overall RPE 6-20 can tentatively be used to assess and form the basis for regulating upper-body exercise at a moderate to vigorous intensity in adults with chronic SCI who have high fitness levels, have been familiarized with the measure and are prompted with the scale during exercise (Supplement 4).

Conclusions

This systematic review showed that there is currently a lack of robust evidence regarding the reliable and valid use of subjective measures to assess aerobic exercise intensity in adults with SCI. Health care professionals and scientists need to be aware of this limited evidence base, which currently prohibits a strong clinical recommendation towards use of these subjective measures. Still, it seems appropriate to provide a tentative, conditional recommendation for the use of overall RPE 6-20 to assess exercise intensity, dependent on participants' fitness levels as well as the exercise intensity and modality used.

Figure and Table legends

Figure 1. Flow diagram of studies through the different phases of the review.

Table 1. Eligible study designs to assess test-retest reliability or criterion validity of subjective measures for assessing aerobic exercise intensity.

Table 2. Benchmarks for acceptable, unacceptable or inconclusive test-retest reliability and criterion validity.

Table 3. Data extracted from the eligible studies regarding test-retest reliability and criterion validity (alphabetically ordered).

Table 4 Synthesis and appraisal of evidence regarding criterion validity: GRADE assessments and Conclusions.

REFERENCES

1. WHO. Spinal cord injury - Fact sheet of World Health Organization. 2013. Available at: http://www.who.int/mediacentre/factsheets/fs384/en/. Accessed July 31, 2017.

2. Martin Ginis KA, Ma JK, Latimer-Cheung AE, Rimmer JH. A systematic review of review articles addressing factors related to physical activity participation among children and adults with physical disabilities. Health psychology review 2016; 10(4):478-494.

3. Kirshblum SC, Burns SP, Biering-Sorensen F, et al. International standards for neurological classification of spinal cord injury (revised 2011). J Spinal Cord Med 2011; 34(6):535-546.

4. Haisma JA, van der Woude LH, Stam HJ, Bergen MP, Sluis TA, Bussmann JB. Physical capacity in wheelchair-dependent persons with a spinal cord injury: a critical review of the literature. Spinal Cord 2006; 44(11):642-652.

5. van den Berg-Emons RJ, Bussmann JB, Stam HJ. Accelerometry-based activity spectrum in persons with chronic physical conditions. Arch Phys Med Rehabil 2010; 91(12):1856-1861.

6. Martin Ginis KA, Latimer AE, Arbour-Nicitopoulos KP, et al. Leisure time physical activity in a population-based sample of people with spinal cord injury part I: demographic and injury-related correlates. Arch Phys Med Rehabil 2010; 91(5):722-728.

7. Libin A, Tinsley EA, Nash MS, et al. Cardiometabolic risk clustering in spinal cord injury: results of exploratory factor analysis. Top Spinal Cord Inj Rehabil 2013; 19(3):183-194.

8. Cragg JJ, Noonan VK, Krassioukov A, Borisoff J. Cardiovascular disease and spinal cord injury: results from a national population health survey. Neurology 2013; 81(8):723-728.

9. Myers J, Lee M, Kiratli J. Cardiovascular disease in spinal cord injury: an overview of prevalence, risk, evaluation, and management. Am J Phys Med Rehabil 2007; 86(2):142-152.

10. Groah SL, Nash MS, Ward EA, et al. Cardiometabolic risk in community-dwelling persons with chronic spinal cord injury. Journal of cardiopulmonary rehabilitation and prevention 2011; 31(2):73-80.

11. Martin Ginis KA, Hicks AL, Latimer AE, et al. The development of evidence-informed physical activity guidelines for adults with spinal cord injury. Spinal Cord 2011; 49(11):1088-1096.

12. Hicks AL, Martin Ginis KA, Pelletier CA, Ditor DS, Foulon B, Wolfe DL. The effects of exercise training on physical capacity, strength, body composition and functional performance among adults with spinal cord injury: a systematic review. Spinal Cord 2011; 49(6):1103-1127.

13. van Koppenhagen CF, Post MW, de Groot S, et al. The longitudinal relationship between wheelchair exercise capacity and life satisfaction after spinal cord injury: a cohort study in the Netherlands. Life satisfaction and wheelchair exercise capacity in the first years after spinal cord injury 2013:101.

14. van Velzen JM, van Leeuwen CM, de Groot S, van der Woude LH, Faber WX, Post MW. Return to work five years after spinal cord injury inpatient rehabilitation: is it related to wheelchair capacity at discharge? J Rehabil Med 2012; 44(1):73-79.

15. de Groot S, Dallmeijer AJ, Post MW, Angenot EL, van den Berg-Emons RJ, van der Woude LH. Prospective analysis of lipid profiles in persons with a spinal cord injury during and 1 year after inpatient rehabilitation. Arch Phys Med Rehabil 2008; 89(3):531-537.

16. Valent LJ, Dallmeijer AJ, Houdijk H, et al. The individual relationship between heart rate and oxygen uptake in people with a tetraplegia during exercise. Spinal Cord 2007; 45(1):104-111.

17. Leicht CA, Bishop NC, Goosey-Tolfrey VL. Submaximal exercise responses in tetraplegic, paraplegic and non spinal cord injured elite wheelchair athletes. Scand J Med Sci Sports 2012; 22(6):729-736.

18. Krassioukov A. Autonomic function following cervical spinal cord injury. Respiratory Physiology & Neurobiology 2009; 169(2):157-164.

19. Hampson DB, St Clair Gibson A, Lambert MI, Noakes TD. The influence of sensory cues on the perception of exertion during exercise and central regulation of exercise performance. Sports Med 2001; 31(13):935-952.

20. Chen MJ, Fan X, Moe ST. Criterion-related validity of the Borg ratings of perceived exertion scale in healthy individuals: a meta-analysis. J Sports Sci 2002; 20(11):873-899.

21. Borg G. Borg's perceived exertion and pain scales. Champaign, IL, US: Human Kinetics; 1998.

22. Paulson TA, Bishop NC, Leicht CA, Goosey-Tolfrey VL. Perceived exertion as a tool to self-regulate exercise in individuals with tetraplegia. Eur J Appl Physiol 2013; 113(1):201-209.

23. Pandolf KB, Billings DS, Drolet LL, Pimental NA, Sawka MN. Differential ratings of perceived exertion and various physiological responses during prolonged upper and lower body exercise. Eur J Appl Physiol Occup Physiol 1984; 53(1):5-11.

24. Eston RG, Brodie DA. Responses to arm and leg ergometry. Br J Sports Med 1986; 20(1):4-6.

25. Kressler J, Nash MS, Burns PA, Field-Fote EC. Metabolic responses to 4 different body weight-supported locomotor training approaches in persons with incomplete spinal cord injury. Arch Phys Med Rehabil 2013; 94(8):1436-1442.

26. Carty A, McCormack K, Coughlan GF, Crowe L, Caulfield B. Increased aerobic fitness after neuromuscular electrical stimulation training in adults with spinal cord injury. Arch Phys Med Rehabil 2012; 93(5):790-795.

27. Giangregorio L, Craven C, Richards K, et al. A randomized trial of functional electrical stimulation for walking in incomplete spinal cord injury: effects on body composition. J Spinal Cord Med 2012; 35(5):351-360.

28. Stevens SL, Caputo JL, Fuller DK, Morgan DW. Effects of underwater treadmill training on leg strength, balance, and walking performance in adults with incomplete spinal cord injury. J Spinal Cord Med 2015; 38(1):91-101.

29. Valent L, Dallmeijer A, Houdijk H, Slootman HJ, Janssen TW, Van Der Woude LH. Effects of hand cycle training on wheelchair capacity during clinical rehabilitation in persons with a spinal cord injury. Disabil Rehabil 2010; 32(26):2191-2200.

30. Kim DI, Lee H, Lee BS, Kim J, Jeon JY. Effects of a 6-Week Indoor Hand-Bike Exercise Program on Health and Fitness Levels in People With Spinal Cord Injury: A Randomized Controlled Trial Study. Arch Phys Med Rehabil 2015; 96(11):2033-2040.e2031.

31. van der Scheer JW, de Groot S, Tepper M, Faber W, Veeger DH, van der Woude LH. Low-intensity wheelchair training in inactive people with long-term spinal cord injury: A randomized controlled trial on fitness, wheelchair skill performance and physical activity levels. J Rehabil Med 2016; 48(1):33-42.

32. Totosy de Zepetnek JO, Pelletier CA, Hicks AL, MacDonald MJ. Following the Physical Activity Guidelines for Adults With Spinal Cord Injury for 16 Weeks Does Not Improve Vascular Health: A Randomized Controlled Trial. Arch Phys Med Rehabil 2015; 96(9):1566-1575.

33. Nooijen CF, van den Brand IL, Ter Horst P, et al. Feasibility of Handcycle Training During Inpatient Rehabilitation in Persons With Spinal Cord Injury. Arch Phys Med Rehabil 2015; 96(9):1654-1657.

34. Bakkum AJ, Paulson TA, Bishop NC, et al. Effects of hybrid cycle and handcycle exercise on cardiovascular disease risk factors in people with spinal cord injury: A randomized controlled trial. J Rehabil Med 2015; 47(6):523-530.

35. Pelletier CA, Totosy de Zepetnek JO, MacDonald MJ, Hicks AL. A 16-week randomized controlled trial evaluating the physical activity guidelines for adults with spinal cord injury. Spinal Cord 2015; 53(5):363-367.

36. Valent LJ, Dallmeijer AJ, Houdijk H, et al. Effects of hand cycle training on physical capacity in individuals with tetraplegia: a clinical trial. Phys Ther 2009; 89(10):1051-1060.

37. Hicks AL, Martin KA, Ditor DS, et al. Long-term exercise training in persons with spinal cord injury: effects on strength, arm ergometry performance and psychological well-being. Spinal Cord 2003; 41(1):34-43.

38. Paulson TA, Bishop NC, Eston RG, Goosey-Tolfrey VL. Differentiated perceived exertion and self-regulated wheelchair exercise. Arch Phys Med Rehabil 2013; 94(11):2269-2276.

39. Lenton JP, Fowler NE, van der Woude L, Goosey-Tolfrey VL. Wheelchair propulsion: effects of experience and push strategy on efficiency and perceived exertion. Appl Physiol Nutr Metab 2008; 33(5):870-879.

40. Martin Ginis KA, Latimer AE, Hicks AL, Craven BC. Development and evaluation of an activity measure for people with spinal cord injury. Med Sci Sports Exerc 2005; 37(7):1099-1111.

41. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. Oxford University Press, USA; 2014.

42. Lamb KL, Eston RG, Corns D. Reliability of ratings of perceived exertion during progressive treadmill exercise. Br J Sports Med 1999; 33(5):336-339.

43. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Bmj 2009; 339:b2700.

44. Morrison A, Polisena J, Husereau D, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care 2012; 28(2):138-144.

45. Morris M, Lamb K, Cotterrell D, Buckley J. Predicting Maximal Oxygen Uptake Via a Perceptually Regulated Exercise Test (PRET). Journal of Exercise Science & Fitness 2009; 7(2):122-128.

46. Whipp BJ, Wasserman K. Oxygen uptake kinetics for various intensities of constant-load work. Journal of Applied Physiology 1972; 33(3):351-356.

47. 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.

48. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012; 21(4):651-657.

49. Kang J, Chaloupka EC, Mastrangelo MA, Donnelly MS, Martz WP, Robertson RJ. Regulating exercise intensity using ratings of perceived exertion during arm and leg ergometry. Eur J Appl Physiol Occup Physiol 1998; 78(3):241-246.

50. WHO. Handbook for guideline development. 2014. Available at: http://apps.who.int/medicinedocs/en/m/abstract/Js22083en/. Accessed Jan 17th, 2017, 2017.

51. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011; 64(4):401-406.

52. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods 2007; 39(2):175-191.

53. Stewart MW, Melton-Rogers SL, Morrison S, Figoni SF. The measurement properties of fitness measures and health status for persons with spinal cord injuries. Arch Phys Med Rehabil 2000; 81(4):394-400.

54. Goosey-Tolfrey V, Lenton J, Goddard J, Oldfield V, Tolfrey K, Eston R. Regulating intensity using perceived exertion in spinal cord-injured participants. Med Sci Sports Exerc 2010; 42(3):608-613.

55. Goosey-Tolfrey VL, Paulson TA, Tolfrey K, Eston RG. Prediction of peak oxygen uptake from differentiated ratings of perceived exertion during wheelchair propulsion in trained wheelchair sportspersons. Eur J Appl Physiol 2014; 114(6):1251-1258.

56. Jacobs PL, Klose KJ, Guest R, Needham-Shropshire B, Broton JG, Green BA. Relationships of oxygen uptake, heart rate, and ratings of perceived exertion in persons with paraplegia during functional neuromuscular stimulation assisted ambulation. Spinal Cord 1997; 35(5):292-298.

57. McCulloch JP, Lorenz DJ, Kloby MA, Love MD, Terson de Paleville PhD DGL. Prediction of Maximal Oxygen Consumption from Rating of Perceived Exertion (RPE) using a Modified Total-body Recumbent Stepper. International Journal of Exercise Science 2015; 8(4):10.

58. Al-Rahamneh HQ, Eston RG. Prediction of peak oxygen consumption from the ratings of perceived exertion during a graded exercise test and ramp exercise test in able-bodied participants and paraplegic persons. Arch Phys Med Rehabil 2011; 92(2):277-283.

59. GRADE. Grading of Recommendations Assessment, Development and Evaluation: Publications. 2017. Available at: http://www.gradeworkinggroup.org/ - pub. Accessed Jan 17th, 2017, 2017.

60. Martin Ginis KA. Letter to the Editor. J Sci Med Sport 2016; 19(8):604.

61. Kilkens OJ, Dallmeijer AJ, Nene AV, Post MW, van der Woude LH. The longitudinal relation between physical capacity and wheelchair skill performance during inpatient rehabilitation of people with spinal cord injury. Arch Phys Med Rehabil 2005; 86(8):1575-1581.

62. Muller G, Odermatt P, Perret C. A new test to improve the training quality of wheelchair racing athletes. Spinal Cord 2004; 42(10):585-590.

63. Harvey LA, Glinsky JV, Bowden JL. The effectiveness of 22 commonly administered physiotherapy interventions for people with spinal cord injury: a systematic review. Spinal Cord 2016; 54(11):914-923.

64. Cowan RE, Ginnity KL, Kressler J, Nash MS, Nash MS. Assessment of the talk test and rating of perceived exertion for exercise intensity prescription in persons with paraplegia. Top Spinal Cord Inj Rehabil 2012; 18(3):212-219.

65. Lewis JE, Nash MS, Hamm LF, Martins SC, Groah SL. The relationship between perceived exertion and physiologic indicators of stress during graded arm exercise in persons with spinal cord injuries. Arch Phys Med Rehabil 2007; 88(9):1205-1211.

66. Andrews JC, Schunemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 2013; 66(7):726-735.

Supplement 1 – Keywords and search strategy for each database

Bibliographic databases and keywords

The following selection of bibliographic databases was searched for studies published from inception until January 1, 2016: Pubmed, PsychINFO (EBSCOhost), SPORTDiscus (EBSCOhost), EMBASE (OVID) and CINAHL (OVID). SCI was represented by keywords such as spinal cord lesion, spine injury or paraplegia, by common non-traumatic causes of SCI (myelitis, myelopathy, spinal cord disease) and by the SCI syndromes that American Spinal Injury Association (ASIA) recognizes (Brown-Sequard, cauda equina, central cord, anterior cord, conus medullaris syndrome).[37] Keywords for subjective exercise intensity were: perceived exertion, perceived effort, perceived intensity, subjective exertion, subjective effort, subjective intensity, perception of exertion, perception of effort and perception of intensity. Each keyword representing SCI was combined with each keyword representing subjective exercise intensity when searching the databases. Language was restricted to English, and expected to have little effect on results.[38] The search strategy for each database is shown below.

Pubmed – Search Strategy

- no filters

((spinal cord*[Text Word] OR spinal cord injur*[Text Word] OR spinal cord disease*[Text Word] OR spinal cord dysfunction*[Text Word] OR spinal cord fracture*[Text Word] OR spinal cord syndrome*[Text Word] OR spinal cord disorder*[Text Word] OR spinal injur*[Text Word] OR spinal disease*[Text Word] OR spinal dysfunction*[Text Word] OR spinal syndrome*[Text Word] OR spinal disorder*[Text Word] OR spinal impairment*[Text Word] OR SCI[Text Word] OR central cord syndrome*[Text Word] OR tetraplegia*[Text Word] OR quadriplegi*[Text Word] OR paraplegi*[Text Word] OR cervical cord*[Text Word] OR Brown-Sequard Syndrome*[Text Word] OR myelitis[Text Word] OR paralys*[Text Word])) AND (perceived exertion*[Text Word] OR perceived effort*[Text Word] OR perceived intensit*[Text Word] OR subjective exertion*[Text Word] OR subjective effort*[Text Word] OR subjective intensit*[Text Word] OR RPE[Text Word])

‘SPORTSDiscus with Full Text’ (via EBSCOhost)

- Box ticked “Also search within the full text of the articles”

<u>TI Title field:</u>
(spinal cord* OR spinal cord injur* OR spinal cord disease* OR spinal cord dysfunction* OR spinal cord fracture* OR spinal cord syndrome* OR spinal cord disorder* OR spinal injur* OR spinal disease* OR spinal dysfunction* OR spinal syndrome* OR spinal disorder* OR spinal impairment* OR SCI OR central cord syndrome* OR tetraplegia* OR quadriplegi* OR paraplegi* OR cervical cord* OR Brown-Sequard Syndrome* OR myelitis OR paralys*) AND (perceived exertion* OR perceived effort* OR perceived intensit* OR subjective exertion* OR subjective effort* OR subjective intensit* OR RPE)

OR

<u>AB Astract field:</u>
(spinal cord* OR spinal cord injur* OR spinal cord disease* OR spinal cord dysfunction* OR spinal cord fracture* OR spinal cord syndrome* OR spinal cord disorder* OR spinal injur* OR spinal disease* OR spinal dysfunction* OR spinal syndrome* OR spinal disorder* OR spinal impairment* OR SCI OR central cord syndrome* OR tetraplegia* OR quadriplegi* OR paraplegi* OR cervical cord* OR Brown-Sequard Syndrome* OR myelitis OR paralys*) AND (perceived exertion* OR perceived effort* OR perceived intensit* OR subjective exertion* OR subjective effort* OR subjective intensit* OR RPE)

OR

<u>KW Keywords</u>
(spinal cord* OR spinal cord injur* OR spinal cord disease* OR spinal cord dysfunction* OR spinal cord fracture* OR spinal cord syndrome* OR spinal cord disorder* OR spinal injur* OR spinal disease* OR spinal dysfunction* OR spinal syndrome* OR spinal disorder* OR spinal impairment* OR SCI OR central cord syndrome* OR tetraplegia* OR quadriplegi* OR paraplegi* OR cervical cord* OR Brown-Sequard Syndrome* OR myelitis OR paralys*) AND (perceived exertion* OR perceived effort* OR perceived intensit* OR subjective exertion* OR subjective effort* OR subjective intensit* OR RPE)

PsycINFO (via EBSCOhost)

- Box ticked "Also search within the full text of the articles"

TI Title field:
(spinal cord* OR spinal cord injur* OR spinal cord disease* OR spinal cord dysfunction* OR spinal cord fracture* OR spinal cord syndrome* OR spinal cord disorder* OR spinal injur* OR spinal disease* OR spinal dysfunction* OR spinal syndrome* OR spinal disorder* OR spinal impairment* OR SCI OR central cord syndrome* OR tetraplegia* OR quadriplegi* OR paraplegi* OR cervical cord* OR Brown-Sequard Syndrome* OR myelitis OR paralys*) AND (perceived exertion* OR perceived effort* OR perceived intensit* OR subjective exertion* OR subjective effort* OR subjective intensit* OR RPE)

OR

AB Astract field:
(spinal cord* OR spinal cord injur* OR spinal cord disease* OR spinal cord dysfunction* OR spinal cord fracture* OR spinal cord syndrome* OR spinal cord disorder* OR spinal injur* OR spinal disease* OR spinal dysfunction* OR spinal syndrome* OR spinal disorder* OR spinal impairment* OR SCI OR central cord syndrome* OR tetraplegia* OR quadriplegi* OR paraplegi* OR cervical cord* OR Brown-Sequard Syndrome* OR myelitis OR paralys*) AND (perceived exertion* OR perceived effort* OR perceived intensit* OR subjective exertion* OR subjective effort* OR subjective intensit* OR RPE)

OR

KW Keywords
(spinal cord* OR spinal cord injur* OR spinal cord disease* OR spinal cord dysfunction* OR spinal cord fracture* OR spinal cord syndrome* OR spinal cord disorder* OR spinal injur* OR spinal disease* OR spinal dysfunction* OR spinal syndrome* OR spinal disorder* OR spinal impairment* OR SCI OR central cord syndrome* OR tetraplegia* OR quadriplegi* OR paraplegi* OR cervical cord* OR Brown-Sequard Syndrome* OR myelitis OR paralys*) AND (perceived exertion* OR perceived effort* OR perceived intensit* OR subjective exertion* OR subjective effort* OR subjective intensit* OR RPE)

OR

TX All Text
(spinal cord* OR spinal cord injur* OR spinal cord disease* OR spinal cord dysfunction* OR spinal cord fracture* OR spinal cord syndrome* OR spinal cord disorder* OR spinal injur* OR spinal disease* OR spinal dysfunction* OR spinal syndrome* OR spinal disorder* OR spinal impairment* OR SCI OR central cord syndrome* OR tetraplegia* OR quadriplegi* OR paraplegi* OR cervical cord* OR Brown-Sequard Syndrome* OR myelitis OR paralys*) AND (perceived exertion* OR perceived effort* OR perceived intensit* OR subjective exertion* OR subjective effort* OR subjective intensit* OR RPE)

STEWART ET AL (2000) - COSMIN CHECKLIST RELIABILITY	Excellent	Good	Fair	Poor
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?*	Adequate sample size			Small sample size
4 Were at least two measurements available?	At least two measurements		Only one measurement	
5 Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6 Was the time interval stated?	Time interval stated	Time interval NOT stated		
7 Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	
9 Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
11 for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described; and/or limits of agreement reported; and/or individual date provided**	ICC calculated but model or formula of the ICC not described or not optimal.	Pearson or Spearman correlation coefficient calculated WITHOUT evidence	No ICC or Pearson or Spearman correlations calculated
12 for dichotomous/nominal/ordinal scores: Was kappa calculated?	N/A			
13 for ordinal scores: Was a weighted kappa calculated?	N/A			
14 for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	N/A			

* for 80% power to detect an ICC≥0.70 or r≥0.80, using a one-tailed test with α=0.05, N≥11 or N≥8 are the minimal sample sizes, respectively (Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods 2007;39(2):175-91).

** Added based on the statistics presented in Table 1