Quality and usability of clinical assessments of static standing and sitting posture

Woldendorp, Kees H; Kleinbergen, Jonas F E; Boonstra, Anne M; de Schipper, Antoine W; Arendzen, J Hans; Reneman, Michiel F

Published in:
Journal of Back and Musculoskeletal Rehabilitation

DOI:
10.3233/BMR-200073
10.3233/BMR-200073

Publication date:
2022

Journal of Back and Musculoskeletal Rehabilitation -1 (2021) 1–16 1
DOI 10.3233/BMR-200073
IOS Press

Review Article

Quality and usability of clinical assessments of static standing and sitting posture: A systematic review

Kees H. Woldendorp[a,*], Jonas F.E. Kleinbergen[a], Anne M. Boonstra[a], Antoine W. de Schipper[b], J. Hans Arendzen[c] and Michiel F. Reneman[d]
[a]*Rehabilitation Expertise Center for Music and Dance, Revalidatie Friesland, Center for Rehabilitation, Beetsterzwaag, The Netherlands*
[b]*Amsterdam University of Applied Sciences, Amsterdam, The Netherlands*
[c]*Department of Rehabilitation Medicine, Leiden University Medical Center, Leiden, The Netherlands*
[d]*Center for Rehabilitation and Department of Rehabilitation, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands*

Received 15 March 2021
Accepted 7 June 2021

Abstract.
BACKGROUND: A validated method to assess sitting and standing posture in a clinical setting is needed to guide diagnosis, treatment and evaluation of these postures. At present, no systematic overview of assessment methods, their clinimetric properties, and usability is available.
OBJECTIVE: The objective of this study was to provide such an overview and to interpret the results for clinical practice.
METHODS: A systematic literature review was performed according to international guidelines. Two independent reviewers assessed risk of bias, clinimetric values of the assessment methods, and their usability. Quality of evidence and strength of recommendations were determined according to the Grading of Recommendations Assessment, Development and Evaluation working group (GRADE).
RESULTS: Out of 27,680 records, 41 eligible studies were included. Thirty-two assessment instruments were identified, clustered into five categories. The methodological quality of 27 (66%) of the articles was moderate to good. Reliability was most frequently studied. Little information was found about validity and none about responsiveness.
CONCLUSIONS: Based on a moderate level of evidence, a tentative recommendation can be made to use a direct visual observation method with global posture recorded by a trained observer applying a rating scale.

Keywords: Occupational medicine, observation, musculoskeletal diseases, posture, reliability, validity

1. Introduction

Among musicians, there is a high prevalence of musculoskeletal complaints [1]. A causal relation is often assumed between 'poor' postures and musculoskeletal complaints in both musicians and non-musicians [2–6]. Identification of asymmetries and other 'abnormali-

*Corresponding author: Kees H. Woldendorp, Rehabilitation Expertise Center for Music and Dance, Revalidatie Friesland, Center for Rehabilitation, PO Box 2, 9244 ZN, Beetsterzwaag, The Netherlands. Tel.: +31 885801295; Fax: +31 885801244; E-mail: k.woldendorp@revalidatie-friesland.nl.

ISSN 1053-8127/$35.00

ties' during static positions is a common procedure in the clinical practice of music medicine, physical therapy, rehabilitation medicine, and occupational medicine [7,8]. It is not clear what a 'poor' or 'risky' posture may be [9–12], nor is there agreement about ways to perform and record observations of sitting and standing poses with a valid, reliable, and clinically usable method.

Reliable information about (working) posture is vital in the diagnosis and treatment of patients with musculoskeletal complaints, to detect potentially risky poses, under the assumption that changing these postures will decrease the problems [3,13–16]. Besides, evaluating the results of therapy and comparing the effects of different treatments to improve posture requires an assessment method that is sensitive to posture changes.

Despite a wide range of literature about aspects of posture assessment, there is little literature about the clinimetric elements of the measurement methods used in daily practice. This is the case for musicians, but also non-musicians. As far as we are aware, there have been few systematic reviews performed following the international guidelines and focusing on assessment methods for global poses – as opposed to specific aspects of posture – that might be suitable for any standing or sitting patient (including musicians). Musicians are singled out here as they are a subgroup of patients with a high prevalence of musculoskeletal complaints [1], and therefore of particular interest in clinical practice.

The most valid and reliable assessment methods, such as multi-camera systems like Optotrak, Vicon, Motion Analysis, or Surface Topography Systems [17–20], are expensive and time-consuming, making large-scale use of this kind of instrumentation in routine clinical settings unrealistic. On the other hand, the widely used assessment method for posture in everyday practice, i.e., the visual observation by a clinician, seems to have low intra- and inter-observer reliability [7,21]. Moreover, visual inspection usually is not performed in a standardized way. Although the training of observers appears to improve the levels of agreement, they are still on a moderate level [20,22].

In order to find a clinically useful and reliable method to assess posture, especially one which can be used in the treatment of juvenile and adult musicians, we performed a systematic review to identify a clinically useful and reliable method to assess static posture. This study aimed to provide an overview of the clinimetric and feasibility properties of the assessment methods for static standing and/or sitting posture in a routine clinical setting and to interpret the findings for clinical practice. Given the limited number of publications focusing on musicians, we have widened the scope of our review to include posture assessment of all kinds of sub-populations in clinical practice.

2. Methods

2.1. Operationalization of the research objective

The terms used in describing the aim of the study were defined as follows:

- 'Assessment method for posture' includes all types of standardized methods by which the posture of a human being can be assessed visually or with the help of, e.g., photography.
- 'Clinimetric properties' (including interpretation, recording, and evaluation) can be assessed in qualitative ways (e.g. 'good/not good', 'risky posture' or 'better/worse') and/or quantitative ways (e.g. 'millimeters/degrees', 'data plotted against reference data for a population' or 'difference in millimeters/degrees').
- 'Suitable for routine practice in a normal clinical setting' means that the instrument is inexpensive, not too space-consuming, transportable, and easy to use without extensive training. Similar requirements apply to the technical aspects. It is essential, for example, that the data obtained should be delivered to health care professionals such as physical therapists, ergonomists, and physicians in a simple format and without delays.
- 'Posture' is the alignment or orientation of the body segments while maintaining a position [23].
  - 'Static' means that the aspects of movement, maintaining balance, or other time-related dynamics are not included.
  - 'Sitting posture,' in the absence of an internationally agreed scientific definition [24], we define this as the situation in which the body is resting on a seat on the buttocks or haunches [25].
  - 'Standing posture' is the position in which a person stands upright with at least one foot on the ground for more than 4 seconds while remaining within a 1 m2 area [26].

2.2. Search

First, electronic medical databases, one trial register, and additional non-electronic channels (grey litera-

Table 1
In- and exclusion criteria

Inclusion criteria (Pool 1-4A)#	Exclusion criteria (Pool 4A)#
1 All languages	1 Articles about assessment instruments for the range of movement or movement
2 Articles about assessment instruments of the observation of $\geqslant 1$ domain of body posture(s)	2 Articles about assessment instruments for body balance
3 Articles about assessment instruments and the assessment method of posture	3 Articles about assessment instruments using a (skills)lab or other complex/ expensive/time spending method not applicable for daily use
4 Articles about assessment instruments based on validation of the instrument (level of evidence A2, B or C*)	4 Articles about assessment instruments measuring over a period of team, with e.g. the mean or number of posture frequencies over time as outcome
	5 Articles about assessment instruments based on interpretation by the authors (systematic review or experts opinion: level of evidence A1 or D*)
	6 Articles about assessment instruments that provided insufficient Information to allow adequate interpretation of outcome measures and results
	Additional exclusion criteria (Pool 4B)#
	7 Non-English papers
	8 Papers published before 1990
	9 Articles about assessment instruments of the observation of < 2 abutting domains of body posture(s) (e.g. head and lower extremities)

#Pool 4A of records: the pool of potential relevant records in the initial search, Pool 4B: final pool of included records, created by additional exclusion criteria for reasons of handling (see text). *CBO-Levels of Evidence (2007): see www.cbo.nl for detailed Information.

ture) were searched for eligible articles. The database search was conducted on December 1, 2017, following the Cochrane guidelines for systematic reviews of diagnostic tests [27]. It covered the electronic databases Cochrane (1940–2017), Medline (PubMed) (1950–2017), Embase (1974–2017), CISDOC (1901–2017), ScienceDirect (1997–2017), Web of Science (1900–2017) and CINAHL (1977–2017). An additional search (using the search terms 'posture AND assessment') was performed in ClinicalTrials.gov in December 2017. 'Grey literature' was searched from December 1, 2014 through December 31, 2017.

Search terms (MeSH and non-MeSH terms) were divided into three domains: 'the instrument' (e.g. method, instrument, technique); 'the goal of the instrument' (e.g. assessing, screening, examining); and 'posture' (e.g. upright position, posture, seated position). We combined individual search terms within each of the domains with the Boolean operator 'OR'. The three domains themselves were combined with the Boolean operator 'AND'. It was anticipated that a massive amount of records would emerge from the databases, given the broad scope of the search terms, and since this is a common feature of systematic reviews about measurement properties [28]. Therefore, to keep the number of records manageable, we added a fourth domain linked to the other three by the Boolean operator 'NOT' to exclude non-relevant titles. Search items were added to this fourth domain until the number of records in the first database (Medline) had dropped to below 15,000 documents (Supplementary Table S1). In the procedure to reduce the number of titles by using 'NOT' + MeSH-terms, and in order not to lose any potentially relevant records, we checked the validity of the procedure by checking whether three of the very appropriated records [11,22,29] found in a previous explorative review (performed by our group) remained in the pool of papers. The search strategies differed slightly for each database (Supplementary Table S1). We searched the references of the relevant papers, as were the reference lists of articles thus identified, and the reference lists of 13 identified review papers about posture assessment [8,11,22,30–39]. 'Grey' literature was collected through various non-electronic channels, i.e., via colleagues, from books, and using a hand search of the journal Medical Problems of Performing Artists from 1986 to 2000. Finally, duplicate articles were removed.

2.3. Selection

The selection procedure is presented in Supplementary Fig. S1. Titles, abstracts, and full texts were screened independently by two reviewers (KHW and JK), in three stages, for their eligibility according to the inclusion and exclusion criteria. This resulted in three pools of potentially relevant records (pools 2, 3, and 4a), as shown in Table 1. Additional exclusion criteria were added after pool 4A had been created because the number of articles was still too large: we excluded articles in other languages than English, articles about assessment instruments limited to the observation of < 2 adjacent domains of body posture(s) (e.g., back and lower extremities), and articles published before

1990. This resulted in the final pool 4b (see Table 1 for inclusion and exclusion criteria).

At each stage of the screening (title, abstract, and full text), the reviewers (KHW and JK) met and resolved disagreement about individual citations through consensus or, if necessary, by consulting a third reviewer (AMB). The two reviewers merged the data into one database and checked whether all data had been entered correctly. The same procedure was performed for entering data in the final tables and figures. The level of agreement between the two reviewers was calculated at all stages using % agreement and Cohen's kappa.

2.4. Missing information

If papers about the clinimetric values of an assessment instrument referred to other publications about the development of that instrument, we included these papers as part of the first paper. If data extraction was not possible, additional information was obtained by contacting the authors listed in the article. Missing information was recorded in the final critical appraisal tables.

2.5. Assessment of quality

The methodological quality of each included study was assessed independently by the two researchers using the Quality Assessment of Diagnostic Accuracy Studies-II (QUADAS-II) checklist [40] and the COnsensus-based Standards for the Selection of health Measurement INstruments (COSMIN) [41,42]. The combined use of these two checklists provided complementary information, despite some overlap: items 3, 4, and 14 in QUADAS-II are identical to H4, B7, and F1, F2, H5, respectively, in COSMIN. Most questions in COSMIN that refer to the presence of restrictions regarding design requirements and statistical methods ask for more details in comparison to the QUADAS-II items. Items 1, 2 and 6–13 of the QUADAS-II are not included in COSMIN.

The QUADAS-II instrument consists of four domains: patient selection, index test, reference standard, and 'flow and timing' (flow of patients through the study and timing of the index tests and reference standard) [40]. We graded the risk of bias in patient selection, index test, reference standard, and flow and timing as high, unclear, or low. The same assessment strategy was used for applicability regarding patient selection, index test, and reference standard.

The items (boxes) of COSMIN [41,42] were used to determine the clinimetric values of the instruments. The correlation coefficient of reliability was interpreted as follows: values > 0.75 as good, those between 0.50 and 0.75 as moderate, and those < 0.5 as poor reliability [43]. Interpretation of the correlation coefficient for criterion and concurrent validity was as follows: values $\geqslant |0.70|$ as strong, between $|0.3|$ and $|0.70|$ as moderate, and values $\leqslant |0.3|$ as weak [38].

We developed a self-constructed customized checklist (Supplementary Table S2) to measure the clinical usability of each measurement instrument for posture. Aspects included in the list were readability of the instructions, comprehensibility, time required to administer the tool, physical requirements (e.g., camera, space, researcher), and the effort involved in interpretation. Each aspect was scored with points ranging from -2 to 2 or -1 to 1. A sum score for each posture measurement instrument was calculated by summing the item scores. Sum scores were calculated for both the clinimetric aspects and clinical usability to enable us to balance the clinical use and scientific support for each measurement instrument.

Results were aggregated and interpreted according to the framework for therapeutic and diagnostic tests developed by the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) working group [44–49]. The framework for therapeutic studies covers five aspects of quality of evidence (study design, inconsistency of results, indirectness of evidence, imprecision, and reporting bias) and four elements of the strength of recommendation (quality of the evidence, uncertainty about the balance between desirable and undesirable effects, uncertainty or variability in values and preferences, and uncertainty about whether the intervention involves extensive use of resources) [45]. Details about categorizing the above five aspects, aggregation of the different scores according to the GRADE framework, and calculation strength of recommendation are provided in Supplement Text 1.

We determined the probability of publication bias by comparing the size of the study sample with the level of the inter-rater reliability values. If smaller studies (< 30 participants) had higher inter-rater reliability values than the more extensive studies ($\geqslant 30$ participants), this could be an indication of publication bias.

We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist of 'items to include when reporting a systematic review' [50] and the checklist 'A Measurement Tool to Assess Systematic Reviews' (AMSTAR) [51] to optimize our reporting of the present review. The study protocol was accepted for registration in the PROSPERO register (no. CRD42017041711).

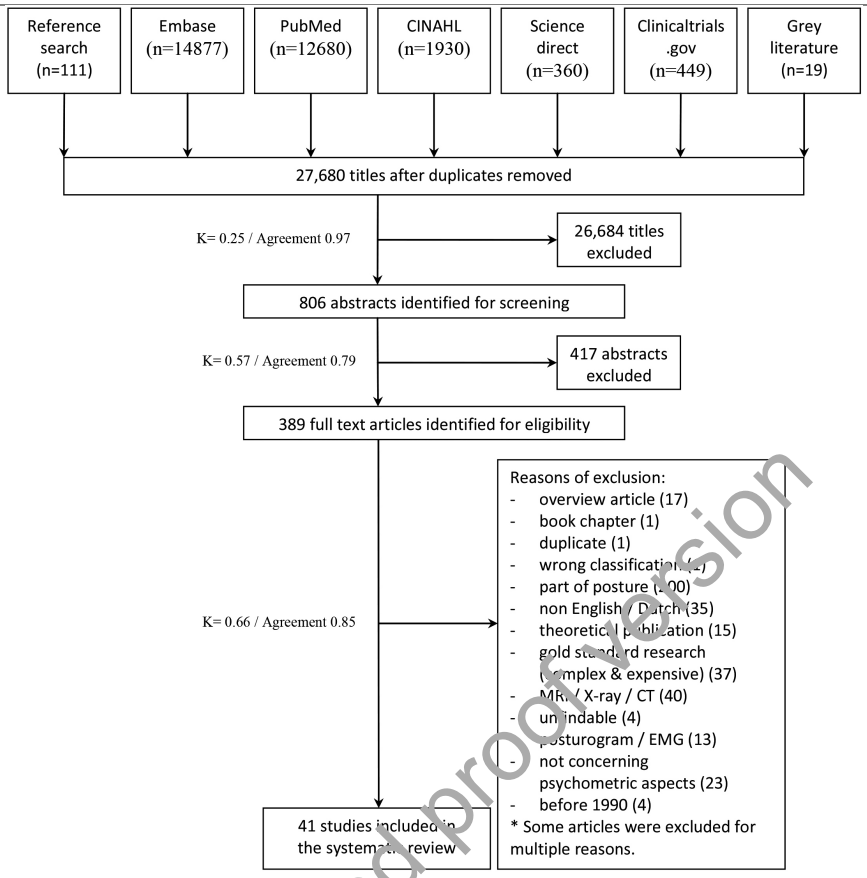

Fig. 1. Flowchart of the screening process.

3. Results

3.1. Search

The first search identified 27,680 papers, 389 of which were retrieved in full text and screened for eligibility. In the end, 41 of these papers we included in the review. Results of the screening and selection process are presented in Fig. 1. For one record, [21] the decision to incorporate was made by the third reviewer (AMB). In the final step of the selection process, the agreement between the two screeners was good ($K = 0.66$) (Fig. 1). Because some studies included more than one method and/or clinimetric value as an outcome parameter, the sum of results can be different from 41. Main reasons for exclusion of full text papers were 200 papers related to the study of only one body part, 35 papers in non-English/Dutch, and 37 papers focused on measurement instruments not available in routine practice (like VICON).

Data relating to standing and sitting postures are presented in Supplement Table 5; twenty-two studies reported data about standing position, five studies about sitting posture, while sixteen studies reported mixed data about both standing and sitting postures.

3.2. Study characteristics

The characteristics of the included studies are presented in Supplementary Table S3. Twenty-nine studies focused on the intra- and inter-rater reliability of an assessment instrument, and six on test-retest reliability. Eleven studies assessed clinimetric aspects concerning the reliability and validity of one assessment instrument. Seven out of the 13 articles about validity concerned concurrent or criterion validity, an item not included in the COSMIN checklist [41]. The eight studies comparing two instruments – neither of which was considered the gold standard – were evaluated using the COSMIN Box (Box H) for criterion validity. We chose one of these two instruments as the reference standard and considered to be the gold standard, though with the qualification of ‘not a good gold standard’.

Studies concerning aspects of validity were incom-

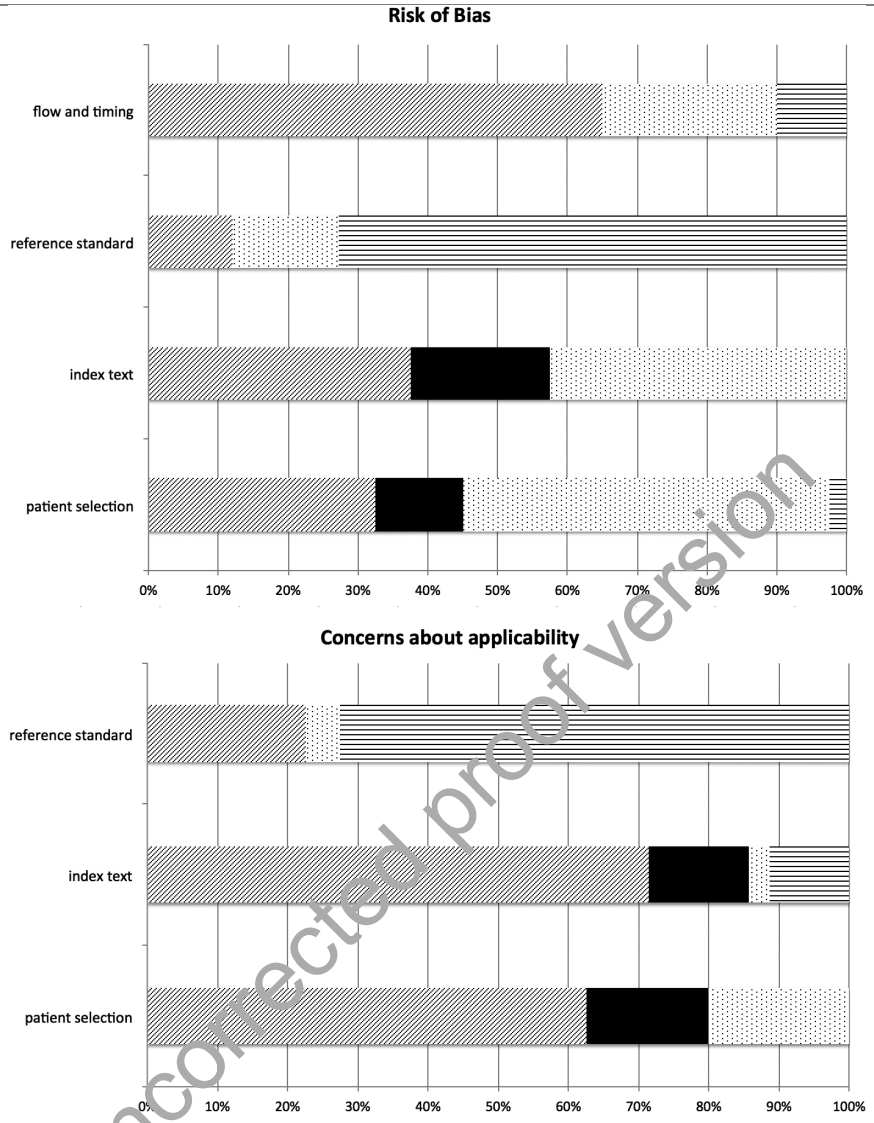

Fig. 2. Graphical display of the QUADAS-II assessments (data from Supplement S4). Risk of bias/applicability: diagonal shading = Low; dotted shading = unclear; solid shading = high; horizontal shading = not applicable.

plete, and the diversity of the validity items was vast. The study settings were mainly work, laboratory, or school. The total study sample consisted of over 2,600 men and women, aged 5–86 years, with a majority of the sample aged between 18 and 40 years. The study sample of 27 studies was the adult working and/or general population, while six studies focused on children (< 18 years) and only two on musicians.

3.3. Study quality

The results of our critical appraisal of the study quality are presented in Supplementary Table S4 and Fig. 2 (QUADAS-II) [40]. The methodological quality of the studies varied considerably (Fig. 2): 11 out of 41 papers (26.8%) had a score of excellent, with a low risk of bias, low concern regarding applicability, or a maximum of one 'unclear' rating for all items scored. Sixteen studies (39.0%) had a score of moderate, with a maximum of three 'unclear' grades, one high risk of bias, or major concern regarding applicability. Thus, 66% of the studies had at least a moderate level of methodological quality. The rest of the papers (34%) had a poor level, with at least two high risks of bias or major concerns about applicability and/or at least four 'unclear' ratings.

Risk of financial conflicts is not listed in the

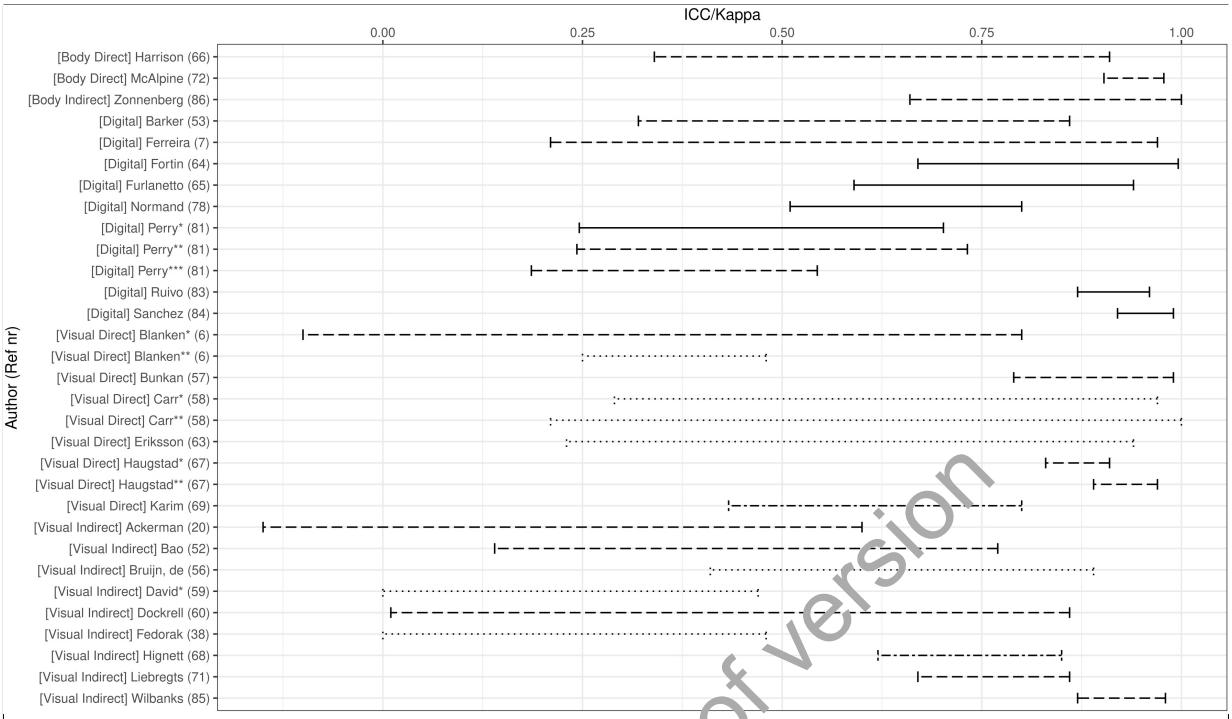

Fig. 3. Inter-observer reliability of the assessment instruments per category. Coefficient type: solid line = ICC; dashed line = ICC (High Quality study); long dashed line = Kappa. Asterisk = multiple instruments in study

QUADAS-II, but is important in view of study quality. From 31 out of 41 papers the authors reported no conflicts of (financial) interest, from nine papers it was not clear if there was a conflict of interest [6,21,54,63,65,69,70,73,77], but one paper [78] mentioned that the first author was paid by the manufacturer (of Posture-Print).

3.4. Characteristics of assessment instruments

We identified a total of 32 assessment instruments (Supplementary Table S5). These were categorized into five groups of assessment methods:

1. Direct body measurement
2. Indirect body measurement (via photograph/video still)
3. Direct visual observation
4. Indirect visual observation (via photograph/video still)
5. Digital measurement: software interpretation of digital 2D-3D photographs/video stills

In 22 (53.7%) of the studies, a continuous scale was predefined for recording the scores obtained. The head and trunk were the most frequently studied body domains (in 38 articles). The upper and lower extremity domains were less often studied, in 31 and 23 studies, respectively, and the least studied was the center of mass domain (in five papers). Twenty-six instruments covered the assessment of three or four body domains, while ten tools assessed two adjacent body domains.

3.5. Clinimetric values of assessment instruments

The clinimetric values of the different assessment instruments for the observation of posture are listed in Supplementary Table S6. For none of the assessment instruments/methods were all items of validity and reliability reported. Two studies reported content or construct validity, while none of the studies reported responsiveness.

Most papers concerned the intra- and inter-rater reliability, with 19 and 29 studies, respectively. We, therefore, decided to use inter-rater reliability to compare the five categories of assessment instruments, to obtain some indication of one of the clinimetric properties of the tools.

Figure 3 shows a wide dispersion of values in all categories of assessment instruments. The nature of these items strongly influences the inter-rater reliability values of some posture assessment items (e.g., reliability

of the assessment of the degree of rotation of a posture domain, or frontal view in comparison to sagittal view). This phenomenon was found for all five categories of observation methods. The 11 studies with the highest methodological quality scores according to QUADAS-II, were distributed across two groups of assessment methods: nine studies concerned a digital and two studies a visual direct approach.

Fourteen studies about reliability used measurement error as an indicator; one concerned an indirect body assessment instrument and three concerned indirect visual assessment, while the remaining ten studies concerned the category of digital assessment instruments. The values for standard error of measurement, minimal detectable difference, or standard deviation were low (0.001–9 mm/0.20–3.8o, with two outliers of 23 mm and 28o). The coefficient of variation varied considerably, with wide confidence intervals [54,64,76,80].

We observed no relevant differences between the clinimetric and usability values of the measurement methods for either standing or sitting postures.

3.6. Clinical usability of the assessment instruments

The clinical usability scores of the different assessment instruments and categories are given in Supplementary Table S7 and graphically presented in Fig. 4. The direct and indirect visual assessment instrument categories had the highest clinical usability scores.

To enable a tradeoff between the inter-rater reliability values and the clinical usability of the assessment instrument groups, we presented these data for each category in Fig. 5.

The ideal assessment instrument should have high clinimetric values as well as excellent clinical usability. The methods that came closest to this ideal were one method in the direct [53] and one in the indirect [69] visual assessment categories. Next to these two studies, one other direct visual way [70] and three visual indirect assessment methods [57,72,87] were identified. Five out of these six methods use a rating scale for recording the visual assessment. The categories of digital and body assessment instruments scored high on inter-observer reliability but lowered on clinical usability.

3.7. Aggregation of results

Table 2 summarizes the findings. The number of studies in each of the five categories was small to moderate (range 2–16). Of the five categories, the highest number of studies concerned digital methods and direct and indirect visual methods. All included studies had a cross-sectional design, so ratings of the level of evidence of all methods were restricted to a maximum of 4 points. A wide range of values was found regarding the evaluation of risk of bias, concerns about applicability, consistency of outcomes, and usability scores.

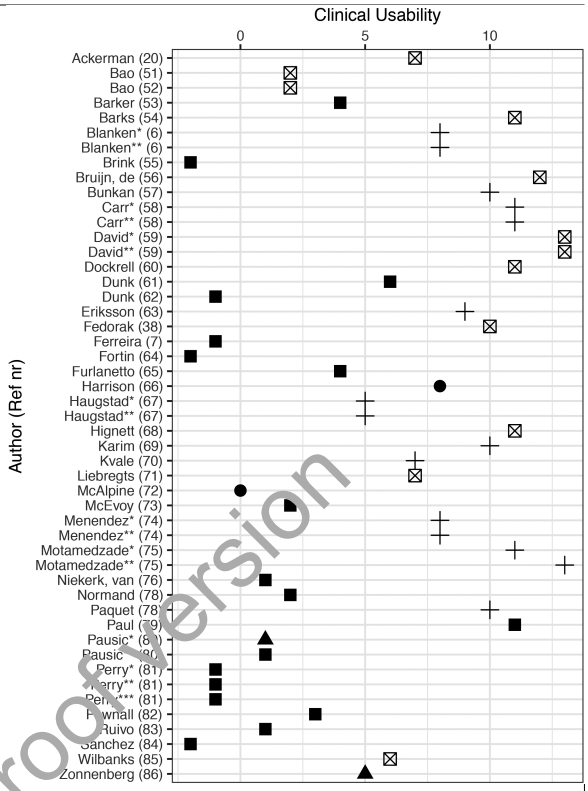

Fig. 4. Clinical usability scores of the assessment instruments Type of instrument: solid circle = body direct; solid triangle = body indirect; solid square = digital; plus sign = visual direct; cross in square sign = visual indirect. Asterisk = multiple instruments in study.

The Visual Direct Measurement methods is the only group of measurement instruments that can be weakly recommended to use as a usable method of measuring global posture in routine practice. For the other groups it is strongly recommended not to use these methods, based on the results of this systematic review.

The risk of publication bias is presented in Supplementary Table S8. We assume that there is a risk of publication bias because there were only two extensive studies ($\geqslant$ 30 participants) identified with high-reliability values, compared to the 8–10 large studies with moderate or low values. The total number of small and comprehensive studies is respectively 17 and 13, but this difference seems not to be an indication for significant publication bias.

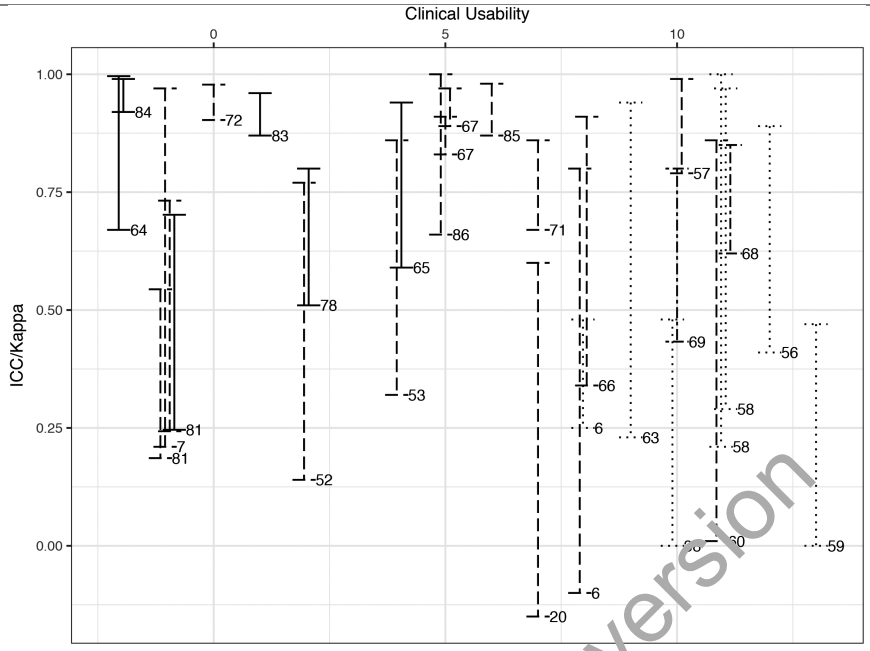

Fig. 5. Inter-rater reliability versus clinical usability of the assessment instruments. Coefficient type: solid line = ICC; dashed line = ICC (High Quality study); long dashed line = Kappa.

4. Discussion

This systematic review aimed to provide an overview of the clinimetric properties of assessment methods for static standing and/or sitting posture in routine clinical settings and to interpret the findings for clinical practice. We identified thirty-two instruments for the clinical assessment of sitting and/or standing position. The tools were divided over five categories: assessment methods using direct body measurements, indirect body measurements (via photographs/video stills), direct visual observations, indirect visual observations, and digital assessment methods (using any form of software to collect information from photographs/video stills).

The following five tentative conclusions were drawn. Firstly, the direct and indirect visual assessment instruments, using a rating scale to record the aspects of posture, seem to have the best combination of inter-rater reliability and usability. Secondly, we found little and incomplete data about validity-related elements and no data about responsiveness. Thirdly, the inter-rater reliability values of some posture assessment items are strongly influenced by the nature of these items (e.g., the reliability of assessing the degree of rotation of a posture domain, or frontal view in comparison to sagittal view). This phenomenon is applied to all five categories of observation methods. Fourthly, the measurement error values (standard error of measurement, minimal detectable difference, and/or standard deviation) are generally low (< 40 or < 9 mm). Fifthly, a weak recommendation (GRADE level B), based on a moderate level of evidence, can be made for clinicians to use the direct visual observation method, using a rating scale, and having it administered by a trained observer. However, the indirect visual observation method has a comparable best combination of inter-rater reliability and usability as the direct visual observation method, but because of other clinimetric aspects we recommend not to use this method in clinical practice. This recommendation is valid for all the other assessment method categories.

The conclusions are partly in line and sometimes conflicting with findings from other reviews [8,11,22, 30–39]. The most similar systematic review [22] about the assessment of biomechanical exposures at work (evaluating both global posture and individual body domains) concluded that none of the observation methods is superior to the others and that global body postures are the most reliable to measure. Our present review comes to different conclusions, as shown by the differences in inter-rater reliability values between the five categories; e.g., the inter-rater reliability values for the direct and indirect body measurement methods are higher than those for the other groups, and there are apparent differences between the five categories in the trade-off between inter-rater reliability and usability

Table 2
Summary of findings and recommendations (GRADE)

Studies	Methodological concerns	Inconsistency of outcomes[2]	Reliability[3]		Validity[5]			Risk of publication Bias[6]	Usability[7] (mean)	Level of recommendation[8]
(Ref. no.)	Risk of bias & Applicability concerns[1]		Intra-rater	Inter-rater	Measurement Error[4]	Criterion & Concurrent V.	Other			
Body measurement direct ($n = 2$)										
Mc Alpine [73]	−2	0	1	1	–	–	–	−1	−1	
Harrison [67]	−1	−1	–	0	–	–	–	0	0	
	−2	**−1**		**1**			–	**−1**	**−1**	**D**
Body measurement indirect ($n = 2$)										
Pausic [81]	−2	–	1	–	0	1	–	1	−1	
Zonnenberg [88]	−1	−1	1	0	–	–	–	−1	0	
	−2	**−1**					–	**0**	**−1**	**D**
Visual measurement direct ($n = 10$)										
Blanken [6]	−1	–	−1	−1	–	–	–	−1	0	
Bunkan [57]	−2	0	–	1	–	–	Internal consistency: 1	−1	1	
Carr [58]	0	–	–	0−1?	–	–	–	0	1	
Eriksson [64]	−1	−1	–	0	–	–	–	−1	1	
Haugstad [68]	−2	0	–	1	–	–	–	0	0	
Karim [70]	0	–	–	0	–	–	–	0	1	
Kvale [71]	0	–	–	–	–	1	Internal consistency: 0	1	0	
Motamedzade [76]	−2	–	–	–	–	0	–	0	1	
Menendez [75]	−1	–	–	–	–	–	–	0	0	
Paquet [79]	−2	–	–	–	–	–	0	−1	0	
	−2	**0**		**0**			–	**0**	**1**	**B**
Visual measurement indirect ($n = 12$)										
Ackermann [21]	−1	−1	−1	−1	–	–	–	0	0	
Bao [52]	−1	–	–	–	–	−1	–	1	−1	
Bao [53]	−2	−1	–	−1	−10.	–	–	−1	−1	
Barks [55]	−1	–	–	−(1)#	–	–	Content/construct validity 1/0	−1/0	1	
David [60]	−2	−1	−1	−1−0?	–	1	0	1	1	
de Bruijn [57]	−1	–	1	1	–	–	–	0	1	
Dockrell [61]	−1	−1	0	0	–	–	–	−1	1	
Fedorak [39]	−2	−1	−1	−1	–	–	–	0	1	
Hignett [69]	−2	–	–	0	–	–	–	1	1	
Liebregts [72]	−1	0	0	0	–	1	–	0	0	
Rodby-Bousquet [84]	−1	1	–	1	−−1	–	Construct validity:1	−1	0	
Wilbanks [87]	−1	1	–	1	–	0	–	−1	0	
	−2	**−1**		**0**			**0**	**0**	**0**	**D**

Table 2, continued

Studies	Methodological concerns	Inconsistency of outcomes[2]	Reliability[3]			Validity[5]		Risk of publication Bias[6]	Usability[7] (mean)	Level of recommendation[8]
(Ref. no.)	Risk of bias & Applicability concerns[1]		Intra-rater	Inter-rater	Measurement Error[4]	Criterion & Concurrent V.	Other			
Digital measurement (n = 16)										
Barker [54]	−1	−1	0	0	–	–	–	−1	0	
Brink [55]	−2	–	–	–	–	1	–	−1	−1	
Dunk [62]	−2	–	–	–	−1	–	Difference: 1	−1	0	
Dunk [63]	−1	–	–	–	–	–	–	−1	−1	
Ferreira [7]	−2	−1	0	0	–	–	–	−1	−1	
Fortin [65]	0	−1	1	1	0.	–	–	0	−1	
Furlanetto [66]	0	0	0	0	0	–	–	0	–	
Mc Evoy [74]	0	–	–	–	–	–	–	0	−1	
Niekerk van [77]	−1	–	–	–	−1	–	–	0	−1	
Normand [78]	0	−1	0	0	0	–	–	0	−1	
Paul [80]	0	–	–	–	–	–	Difference: 1	−1	1	
Pausic [81]	−2	–	1	–	0	–	Construct validity: 1	1	−1	
Perry [82]	0	−1	1	−1−0?	−1.	–	–	−1	−1	
Pownall [77]	−2	−1	–	0	0	–	–	−1	2	
Ruivo [85]	0	1	–	1	0	–	–	1	−1	
Sanchez [86]	0	1	–	1	0	–	–	−1	−1	
	−1	**−1**		**0**			**1**	**−1**	**−1**	D

[1] Interpretation of risk of bias and concerns about applicability (score on QUADAS II): 0 = ⩽ 1 uncertain score/study; −1 = ⩽ 3 uncertain scores or ⩽ 1 high score/study; −2 ⩾ 4 uncertain scores or > 1 high score/study; at group level, the mean value is used for deducting or adding points for this category; the values are rounded. [2] Inconsistency of outcomes (inter-rater reliability): 1 = ⩾ 75% of correlations coefficients ⩾ 0.75/study; 0 = ⩾ 75% of correlations coefficients 0.5 ⩽ correlations coefficients < 0.75/study; −1 = lower percentages of correlation coefficients or lower inter-rater reliability values. [3] Range of scores: 1 = ⩾ 75% of number of values ⩾ 0.75; 0 = 75% of values 0.5 ⩽ values < 0.75; −1 = ⩾ 75% of values < 0.5; at group level, the mean value is used for deducting or adding points for this category, the values are rounded. # = Test-retest value in the absence of inter-rater reliability value. [4] Measurement errors: 0 = ⩽ 10 mm or ⩽ 10 degrees; −1 ⩾ 10 mm or > 10 degrees. [5] Criterion/Concurrent/Construct validity/Internal consistency: 1 = correlation value ⩾ | ± 0.7|; 0 = | ± 0.3| ⩽ correlation value ⩽ | ± 0.7|; −1 = correlation value ⩽ | ± 0.3|; 1 = Differences ⩽ 5 or ⩽ 10 mm; 0 = Differences > 5 or >10 mm. [6] Risk of publication bias: 1 ⩾ 100 participants/study; 0 = 30–100 participants/study; −1 ⩽ 30 participants/study. [7] Usability value (Supplement Table S7): 1 = ⩾ 9 points/study; 0 = 3 ⩽ points/study < 9; −1 ⩽ 3 points/study; at group level the mean value is used for deducting or adding points for this category; the values are rounded. [8] Level of recommendation: The mean sum score (range 1–4) of the initial score at group level and the deduced and added points at group level.

values (with the highest values both being found for the direct and indirect visual observation methods). As far as we know, we are the first that recommend not to use most of the assessment methods for global posture in clinical practice. On the other hand, the wide range of values we found means that, like Takala et al. [22], we can only draw tentative conclusions.

There are several reasons why we need to be careful in drawing too-firm conclusions. The main reason is that for most assessment methods, it is only the reliability that has been thoroughly studied, while the other clinimetric properties are unknown. Another reason is that it is not known to what extent the clinimetric features of the instruments designed to assess the posture of one body domain are comparable with those of devices intended for the assessment of all of these body domains or the global body posture. Our full-text screening excluded 200 papers about assessing the pose of a single body domain. The systematic review by Takala et al. [22] included both single body domain and global posture observation studies. Still, it is not clear how many articles were identified in their paper (in the absence of a flow diagram of the screening procedure). Hence, comparing the outcomes of our study with those of the study by Takala et al. [22] is difficult.

Fortin et al. [11] concluded in their narrative review that the quantitative assessment of global posture is performed most accurately and rapidly by measuring body angles from photographs. This conclusion might be based on their studies of single body domain. The results of our review (about global posture) are not in agreement with this. We found that the digital assessment methods are more suitable for this goal, especially with the advent in recent years of a wide range of new posture assessment apps and photogrammetry software (sometimes freely available on the internet) [36,37]. These are promising assessment methods that might be expected to yield high clinimetric and clinical usability values shortly. These new methods have, however, not yet been tested in validation studies. Moreover, the application of photogrammetry in postural evaluation is directly dependent on both the collection procedures and the mathematical methods used to provide measurements. In line with Fortin et al. [11] and Furlanetto et al. [33], we found that the used postural evaluation software varies significantly among the studies, with often no explanation about the methods used to generate the results. Besides, the software is often not accessible [33]. This lack of data makes it difficult to interpret data synthesis rules within these 'black boxes'.

In the studies included in our review, the measurement error values at the participant group level per study were low, but the confidence intervals were wide. This wideness is due to a combination of variations in marker placement, differences in parameter definitions, body position, perspective error (due to camera position), and especially biological variability (particularly among children due to anthropometric and motor control immaturity). No conclusion can, therefore, be drawn about the ecological validity or the interpretation of these values for individuals in a clinical setting.

The major strengths of our study are a large number of screened and included records, the fact that our conclusions are based on papers of which the majority had moderate to good methodological quality, and the fact that all procedures as much as possible followed the international standards for performing (Cochrane, PRISMA) [27,50] and reporting (AMSTAR) [51] systematic reviews, the systematic and explicit approach (GRADE) [44–49] we judged the quality of evidence and explicit recommendation for clinicians.

Potential limitations of the study are, in theory, a risk of selection bias of articles and a possible bias in the process of interpreting and aggregating the findings. The risk of selection bias is especially relevant for systemic reviews with a broad topic, resulting in a large amount of papers as search result [28]. This is an inevitable consequence of the inclusion of terms such as posture, validity and reliability. Consulted experts in this field had no additional suggestions for minimizing this source of bias. Main reason for exclusion of full text papers were papers related to the study of only one body part. For future reviews, we suggest to analyze these excluded papers, arranged per body domain. The sum of these domain outcomes might be different from global posture.

The risk of selection bias due to publication bias can also be assumed to be relevant. However, the total number of small and large studies didn't differ that much, but because of the low number of extensive studies assigned to the high class of correlation coefficients for the outcomes [43]. There were many choices to be made during the process of interpretation and aggregation, and each of these options required weighing the evidence. The guidelines offer no solution to this problem. An explicit description of the arguments for our choices is provided in this article as much as possible.

Another limitation of the study is that we based the usability values of the assessment instruments on a self-constructed scoring list. We are aware of the subjective nature of this list. As far as we know, there is no objective way to score clinical usability. Before the start of the study, we asked several clinicians to review the scor-

ing list. Most of the discussions and subjective views concerned the item of 'cost price of the assessment instrument'. There were different opinions about the criteria for the various intervals, e.g., depending on the clinical setting (in that the estimation and acceptability of the costs for an instrument used by many therapists in rehabilitation centers differ from those of the same device used by a single therapist in a peripheral physical therapy practice). The cost-price item was also difficult to divide into classes, as little information about it was presented in the papers we reviewed. The bias due to this uncertainty might be that the actual clinical usability scores may be one or two points higher, especially in the indirect assessment categories and the digital category (as these methods use technical support in some form).

Another potential source of bias might be a conflict of financial interest. Although we assume that for the majority of the nine papers without any statement in this respect [6,21,54,63,65,69,70,73,77] this seemed not a major source of bias, it might be a possibility, especially in the group of Digital Measurement Instruments [78]. Manufacturers of products from the latter group have a direct interest in excellent clinimetric outcomes.

Presenting separate data for the sitting or standing posture was only partly possible in our review. Several studies showed combined data for both poses, while in others, it was unclear whether either postures or just one had been included. The consequences of this omission are small; however, as the clinimetric and usability values of the assessment methods are similar for both postures.

It is not clear to what extent the conclusions of our review are generalizable to subpopulations; we found insufficient papers about, e.g. musicians, age groups, patients versus healthy people. What little information we could retrieve from the studies does not appear to show relevant differences, except that a lower level of reliability has to be taken into account with younger children, as their balance maintenance is less mature than that in adults [74].

In line with Takala et al. [22] and Furlanetto et al. [33], we support that selecting a clinical assessment method for posture should be based on the clinician's purpose. Based on our review, it seems best to recommend for the direct visual assessment method, as these provide the best combination of clinimetric and usability values. However, these instruments are less appropriate for the quantification and evaluation of posture and are less responsive to change.

The direct visual observation method is best for situations where a (quick) qualitative observation of pose is required, and/or an estimated quantitative and/or qualitative evaluation of posture (e.g., classification in a rating scale with three classes).

Given the near absence of studies evaluating the construct validity or predictive validity of assessment instruments for static sitting and standing positions, we recommend clinicians to use with caution any possible assessment method for the detecting of postures at risk for musculoskeletal complaints. We also found little information about the criterion validity aspects. In other words, assumptions about what is relevant in assessing and judging static postures – in terms of carrying a risk of musculoskeletal complaints – should be critically reconsidered. In terms of the GRADE framework, the level of recommendation for most diagnostic instruments is often low, because data about these aspects are scarce [48]. Based on these arguments, we tentatively recommend the use of the standardized direct visual observation method for the assessment of static posture. The results of our review do not support the use of other tools in clinical practice.

5. Conclusion

Based on a moderate level of evidence, a weak recommendation can be made for using the direct visual assessment method (with posture recorded as rating scores by a trained observer) to assess sitting and/or standing pose in daily clinical practice. Little and incomplete information was found about validity-related aspects and no data about responsiveness. For all five categories of observation methods, the inter-rater reliability values of some posture assessment items are strongly influenced by the nature of these items.

Acknowledgments

The authors would like to thank Prof. Henrica CW de Vet, Ph.D. and Ms. Wieneke Mokkink, Ph.D. for their theoretical support in the construction phase of the study, and Mr. Guus van den Brekel and Mrs. Sjoukje van der Werf for their technical advice and support in the search strategy phase of the study.

Author contributions

All authors contributed to the study concept and design, the analysis of the results, and the writing of the

manuscript. KHW and JFEK performed the search, selection, and data-extraction, with AMB as consultant. AWS designed the tables and figures. KHW wrote the draft versions of the manuscript. All authors provided critical feedback on the draft versions of the manuscript, and approved the final version.

Conflict of interest

The authors report no declarations of interest.

Supplementary data

The supplementary files are available to download from http://dx.doi.org/10.3233/BMR-200073.

References

[1] Kok LM, Huisstede BM, Voorn VM, Schoones JW, Nelissen RG. The occurrence of Musculoskeletal complaints among professional musicians: a systematic review. Int Arch Occup Environ Health. 2016; 89(3): 373-396.

[2] Cailliet R. Abnormalities of the sitting postures of musicians. Med Probl Perform Art. 1990; 5: 131-135.

[3] National Research Council and the Institute of Medicine. Musculoskeletal disorders and the workplace: low back and upper extremities. Panel on Musculoskeletal Disorders and the Workplace. Commission on behavioral and Social Sciences and Education. Washington, DC: National Academy Press. 2001.

[4] Pascarelli EF, Hsu YP. Understanding work-related upper extremity disorders: clinical findings in 485 computer users, musicians, and others. J Occup Rehabil. 2001; 11: 1-21.

[5] Picavet HS, Schouten JS. Musculoskeletal pain in The Netherlands: prevalences, consequences and risk groups, the DMC(3)-study. Pain. 2003; 102: 167-178.

[6] Windt van der DA, Thomas E, Pope DP, et al. Occupational risk factors for shoulder pain: a systematic review. Occup Environ Med. 2000; 57: 433-442.

[7] Blanken WCG, Rijst van der H, Mulder PGH, Eijsden van Besselink MDF, Lankhorst GJ. Interobserver and intraobserver reliability of postural examination. Med Probl Perform Art. 1991; 6: 93-97.

[8] Ferreira EAG, Duarte M, Maldouado EP, Burke TN, Marques AP. Postural assessment software (PAS/SAPO): validation and reliability. Clinics. 2010; 65: 675-681.

[9] Abréu-Ramos AM, Brandfonbrener AG. Lifetime prevalence of upper-body musculoskeletal problems in a professional-level symphony orchestra: age, gender, and instrument-specific results. Med Probl Perform Art. 2007; 22: 97-104.

[10] Brandfonbrener AG., History of playing-related pain in 330 university freshman music students. Med Probl Perform Art. 2009; 24: 30-36.

[11] Fortin C, Ehrmann-Feldman D, Cheriet F, Labelle H. Clinical methods for quantifying body segment posture: a literature review. Given and Rehab. 2011; 33: 367-383.

[12] Woldendorp KH, Tijsma A, Boonstra AM, Arendzen JH, Reneman MF. No association between posture and musculoskeletal complaints in a professional bassist sample. EJP. 2016; 20: 399-407.

[13] Aghilinejad M, Azar NS, Ghasemi MS, Dehghan N, Mokamelkhah EK. An ergonomic intervention to reduce musculoskeletal discomfort among semiconductor assembly workers. Work. 2016; 54: 445-450.

[14] Kilroy N, Dockrell S. Ergonomic intervention: its effect on working posture and musculoskeletal symptoms in female biomedical scientists. Br J Biomed Sci. 2000; 57: 199-206.

[15] Pillastrini P, Mugnai R, Bertozzi L, et al. Effectiveness of an ergonomic intervention on work-related posture and low back pain in video display terminal operators: a 3 year cross-over trial. Appl Ergon. 2010; 4: 436-443.

[16] Rosario JL. Relief from back pain through postural adjustment: a controlled clinical trial of the immediate effects of muscular chain therapy (MCT). Int J Ther Massage Bodywork. 2014; 7: 2-6.

[17] Goldberg CJ, Kaliszer M, Moore DP, Fogarty EE, Dowling FE. Surface topography, Cobb angles, and cosmetic change in scoliosis. Spine. 2001; 26: E55-E63.

[18] Nault ML, Allard P, Hinse S, et al. Relations between standing stability and body posture parameters in adolescent idiopathic scoliosis. Spine. 2002; 27: 1911-1917.

[19] Pazos V, Cheriet F, Song L, Labelle H, Dansereau J. Accuracy assessment of human trunk surface 3D reconstructions from an optical digitizing system. Med Biol Eng Comput. 2005; 43: 11-15.

[20] Zabjek KF, Leroux MA, Rivard CH, Prince F. Evaluation of segmental postural characteristics during quiet standing in control and idiopathic scoliosis patients. Clin Biomech. 2005; 20: 483-490.

[21] Ackermann BJ, Adams R. Interobserver reliability of general practice physiotherapists in rating aspects of the movement patterns of skilled violinists. Med Probl Perform Art. 2004; 19: 3-11.

[22] Takala EP, Pehkonen I, Forsman M, Hansson GA, Mathiassen SE, Neumann WP, et al. Systematic evaluation of observational methods assessing biomechanical exposures at work. Scan J Work, Environ & Health. 2010; 36(1): 3-24.

[23] Raine S, Twomey L. Attributes and qualities of human posture and their relationship to dysfunction or musculoskeletal pain. Crit Rev Phys and Rehabil Med. 1994; 6: 409-437.

[24] Yates T, Wilmot EG, Davies MJ, et al. Sedentary behavior: what's in a definition? Am J Prev Med. 2011; 40: e33-e34.

[25] Dictionary. (n.d.). Retrieved August 18, 2016 from https://wwwmerriam-webster.com/.dictionary/dictionary/sit.

[26] Mens en Werk, Arbo catalogus. Retrieved August 18, 2016 from http://wwwii-mensen.werk.nl/Upload/Arbocatalogus.

[27] The Cochrane Collaboration. Handbook for DTA reviews. Retrieved November 12, 2014 from http://srdta.cochrane.org/handbook-dta-reviews.

[28] Vet de HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine. 3rd printing. Cambridge: Cambridge University Press; 2014.

[29] Fedorak C, Ashworth TS, Marshall J, Paull H. Reliability of the visual assessment of cervical and lumbar lordosis: how good are we? Spine. 2003; 28: 1857-1859.

[30] Baxter ML, Milosavljevic S, Ribeiro DC, Hendrick P, McBride D. Psychometric properties of visually based clinical screening tests for risk of overuse injury. Phys Ther Rev. 2014; 19: 213-219.

[31] Burt S, Punnett L. Evaluation of interrater reliability for posture observations in a field study. Appl Erg. 1999; 30: 121-135.
[32] Field D, Livingstone R. Clinical tools that measure sitting posture, seated postural control or functional abilities in children with motor impairments: a systematic review. Clin Rehab. 2013; 27(11): 994-1004.
[33] Furlanetto TS, Sedrez JA, Tarragô Condotti C, Fagundes Loss J. Photogrammetry as a tool for postural evaluation of the spine: a systematic review. World J Orthop. 2016; 7(2): 136-148.
[34] Gallagher S. Physical limitations and musculoskeletal complaints associated with work in unusual or restricted postures: a literature review. J Safety Research. 2005; 36: 51-61.
[35] Kilbom A. Assessment of physical exposure Ph.D. work-related musculoskeletal disorders: what information can be obtained from systematic observations? Scan J Work Environ Health. 1994; 20 Special issue: 30-45.
[36] Krawczky B, Pacheco AG, Mainenti MRM. A systematic review of the angular values obtained by computerized photogrammetry in sagittal plane: a proposal for reference values. J Manipulative Physiol Ther. 2014; 37: 269-275.
[37] Linn JM. Using digital image processing for the assessment of postural changes and movement patterns in bodywork clients. J Bodywork Mov Ther. 2001; 5: 11-20.
[38] Singla D, Veqar Z. Methods of postural assessment for sportspersons. J Clin Diagn Research. 2014; 8: LE01-LE04.
[39] Wongthree WY, Wong MS, Lo KH. Clinical applications of sensors for human posture and movement analysis: a review. Prosthet Orthot Int. 2007; 31: 62-75.
[40] Whiting PF, Rutjes AWS, Westwood ME, et al. A revised instrument for the quality assessment of diagnostic accuracy studies. Ann InternBehavioral Med. 2011; 155: 529-536.
[41] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010; 63: 737-745.
[42] Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, Vet de HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012; 21: 651-657.
[43] Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 2nd ed. Upper Saddle River: J. Alexander; 2000.
[44] Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. The GRADE Working Group. BMC Health Serv Res. 2004; 4: 38. doi: 10.1186/1472-6963-4-38.
[45] Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008; 336: 924-926.
[46] Guyatt GH, Oxman AD, Vist GE, et al. GRADE: what is "quality of evidence" and why is it important to clinicians. BMJ. 2008; 336: 995-998.
[47] Guyatt GH, Oxman AD, Vist GE, et al. GRADE: going from evidence to recommendations. BMJ. 2008; 336: 1049-1051.
[48] Guyatt GH, Oxman AD, Vist GE, et al. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ. 2008; 336: 1106-1110.
[49] Guyatt GH, Oxman AD, Vist GE, et al. GRADE: incorporating considerations of resources use into grading recommendations. BMJ. 2008; 336: 1170-1173.
[50] Moher D, Librati A, Tetzlaff J, Altman DG, and the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009; 151: 264-269.
[51] Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement instrument to access the methodological quality of systematic reviews. BMC Med Research Methodol. 2007; 7: 10. doi: 10.1186/1471-2288-7-10.
[52] Bao S, Howard N, Spielholz P, Silverstein B. Two posture analysis approaches and their application in a modified rapid upper limb assessment evaluation. Ergonomics. 2007; 50: 2118-2136.
[53] Bao S, Howard N, Spielholz P, Silverstein B, Polissar N. Interrater reliability of posture observations. Hum Factors. 2009; 51: 292-309.
[54] Barker SP, Tremback-Ball A, Sterner K. Reliability of posture measurements taken with photo analysis. J Women Health Physical Ther. 2006; 30: 2.
[55] Barks L, Luther SL, Brown LM, Schulz B, Bowen ME, Powell-Cope G. Development, and initial validation of the Seated Posture Scale. JRRD. 2015; 52(2): 201-210.
[56] Brink Y, Louw Q, Grimmer K, Schreve K, Westhuizen van der G, Jordaan E. Development of a cost-effective three-dimensional posture analysis tool: validity and reliability. BMC Musculoskelet Disord. 2013; 14: 335.
[57] Bruijn de I, Engel JA, van der Gulden JW. A simple method to evaluate the reliability of OWAS observations. Appl Ergon. 1998; 29: 281-283.
[58] Burkan BH, Moen O, Opjordsmoen S, Ljunggren AE, Friis S. Interrater reliability of the comprehensive body examination. Physioth Theory Pract. 2002; 18: 121-129.
[59] Carr EK, Kenney FD, Wilson-Barnett J, Newham DJ. Interrater reliability of postural observation after stroke. Clin Rehab. 1999; 13: 229-242.
[60] David G, Woods V, Li G, Buckle P. The development of the Quick Exposure Check (QEC) for assessing exposure to risk factors for work-related musculoskeletal disorders. Appl Ergon. 2008; 39: 57-69.
[61] Dockrell S, O'Grady E, Bennett K, et al. An investigation of the reliability of Rapid Upper Limb Assessment (RULA) as a method of assessment of children's computing posture. Appl Ergon. 2012; 43: 632-636.
[62] Dunk NM, Chung YY, Compton DS, Callaghan JP. The reliability of quantifying upright standing postures as a baseline diagnostic clinical tool. J Manipulative Physiol Ther. 2004; 27: 91-96.
[63] Dunk NM, Lalonde J, Callaghan JP. Implications for the use of postural analysis as a clinical diagnostic tool: reliability of quantifying upright standing spinal postures from photographic images. J Manipulative Physiol Ther. 2005; 28: 386-392.
[64] Eriksson EM, Mokhtari M, Pourmotamed L, Holmdahl L, Eriksson H. Inter-rater reliability in a resource-oriented physiotherapeutic examination. Physioth Theory Pract. 2000; 16: 95-103.
[65] Fortin C, Feldman DE, Cheriet F, Gravel D, Gauthier F, Labelle H. Reliability of a quantitative clinical posture assessment tool among persons with idiopathic scoliosis. Physiother. 2012; 98: 64-75.
[66] Furlanetto TS, Candotti CT, Sedrez JA, Noll M, Loss JF. Evaluation of the precision and accuracy of the DIPA software postural assessment protocol. Eur J Physiother. 2017; 19(4): 179-184.
[67] Harrison AL, Barry-Greb T, Wojtowicz G. Clinical measurement of head and shoulder posture variables. J Orthop Sports Phys Ther. 1996; 23: 353-361.
[68] Haugstad GK, Haugstad TS, Kirste U et al. Reliability and

validity of a standardized Mensendieck physiotherapy test (SMT). Physiother Theory Pract. 2006; 22: 189-205.

[69] Hignett S, McAtamney L. Rapid entire body assessment (REBA). Appl Ergon. 2000; 31: 201-205.

[70] Karim A, Millet V, Massie K, Olson S, Morganthaler A. Inter-rater reliability of a musculoskeletal screen as administered to female professional contemporary dancers. Work. 2011; 40: 281-288.

[71] Kvale A, Bunkan BH, Opjordsmoen S, Ljunggren AE, Friis S. Development of the posture domain in the Global Body Examination (GBE). Adv Physiother. 2010; 12: 157-165.

[72] Liebregts J, Sonne M, Potvin JR. Photograph-based ergonomic evaluations using the Rapid Office Strain Assessment (ROSA). Appl Ergon. 2016; 52: 317-324.

[73] McAlpine RT, Betaany-Saltikov JA, Warren JG. Computerized back postural assessment in physiotherapy practice: Intra-rater and inter-rater reliability of the MIDAS system. J Back Musculoskelet Rehabil. 2009; 22: 173-178.

[74] McEvoy MP, Grimmer K. Reliability of upright posture measurements in primary school children. BMC Musculoskelet Disord. 2005; 6: 35. doi: 10.1186/1471-2474-6-35.

[75] Menendez CC, Amick IIIDoi BC, Chang CC, et al. Evaluation of two posture survey instruments for assessing computing postures among college students. Work. 2009; 34: 421-430.

[76] Motamedzade M, Ashuri MR, Golmohammadi R, Mahjub H. Comparison of ergonomic risk assessment outputs from rapid entire body assessment and quick exposure check in an engine oil company. J Res Health Sci. 2011; 11: 26-32.

[77] Niekerk van SM, Louw Q, Vaughan C, Grimmer-Somers K, Schreve K. Photographic measurement of upper-body sitting posture of high school students: a reliability and validity study. BMC Musculoskelet Disord. 2008; 9: 113. doi: 10.1186/1471-2474-9-113.

[78] Normand MC, Descarreaux M, Harrison DD, et al. Three dimensional evaluation of posture in standing with the PosturePrint: an intra- and inter-examiner reliability study. Chiropr Osteopat. 2007; 15: 15. doi: 10.1186/1746-1340-15-15.

[79] Paquet VL, Punnett L, Buchholz B. Validity of fixed-interval observations for postural assessment in construction work. Appl Ergon. 2001; 32: 215-224.

[80] Paul JA, Douwes M. Two-dimensional photographic posture recording and description: a validity study. Appl Ergon. 1993; 24: 83-90.

[81] Pausić J, Pedisić Z, Dizdar D. Reliability of a photographic method for assessing standing posture of elementary school students. J Manipulative Physiol Ther. 2010; 33: 425-431.

[82] Perry M, Smith A, Straker L, Coleman J, O'Sullivan P. Reliability of sagittal photographic spinal posture assessment in adolescents. Adv Physiother. 2008; 10: 66-75.

[83] Pownall PJ, Moran RW, Stewart AM. Consistency of standing and seated posture of asymptomatic male adults over a one-week interval: a digital camera analysis of multiple landmarks. Int J Osteopath Med. 2008; 11: 43-51.

[84] Rodby-Bousquet E, Ágústsson A, Jónsdóttir G, Czuba T, Johansson AC, Hägglund G. Interrater reliability and construct validity of the Posture and Postural Ability Scale in adults with cerebral palsy in supine, prone, sitting and standing positions. Clin Rehab. 2014; 28: 8: 00.

[85] Ruivo RM, Pezarat-Correia P, Carita AI. Intrarater and interrater reliability of photographic measurement of upper-body standing posture of adolescents. J Manipulative Physiol Ther. 2015; 38(): 74-80.

[86] Sánchez MB), Loram I, Darby J, Holmes P, Butler PB. A video-based method to quantify posture of the head and trunk in sitting. Gait Post. 2017; 51: 181-187.

[87] Vidanks S. To determine the reliability of a digital photo-based posture assessment in standing. Arch Phys Med Rehab. 2016; 97(12): e6.

[88] Zonnenberg AJ, Maanen van CJ, Elvers JW, Oostendorp RA. Intra/interrater reliability of measurements on body posture photographs. Cranio. 1996; 14: 326-331.