Manual Therapy 19 (2014) 10–17

Contents lists available at ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/math

Systematic review

Reliability and validity of non-radiographic methods of thoracic kyphosis measurement: A systematic review

Eva Barrett [a,*], Karen McCreesh [a], Jeremy Lewis [b,c]

[a] Department of Clinical Therapies, Faculty of Education and Health Sciences, University of Limerick, Limerick, Ireland
[b] Musculoskeletal Services, Health at the Stowe, Central London Community Healthcare, NHS Trust, 260 Harrow Road, London W2 5ES, UK
[c] Department of Allied Health Professions and Midwifery, School of Health and Social Work, Wright Building, College Lane Campus, University of Hertfordshire, Hatfield, AL10 9AB, Hertfordshire, UK

ARTICLE INFO

Article history:
Received 7 January 2013
Received in revised form
29 April 2013
Accepted 16 September 2013

Keywords:
Reliability
Validity
Thoracic kyphosis
Measurement

ABSTRACT

Background: A wide array of instruments are available for non-invasive thoracic kyphosis measurement. Guidelines for selecting outcome measures for use in clinical and research practice recommend that properties such as validity and reliability are considered. This systematic review reports on the reliability and validity of non-invasive methods for measuring thoracic kyphosis.

Methods: A systematic search of 11 electronic databases located studies assessing reliability and/or validity of non-invasive thoracic kyphosis measurement techniques. Two independent reviewers used a critical appraisal tool to assess the quality of retrieved studies. Data was extracted by the primary reviewer. The results were synthesized qualitatively using a level of evidence approach.

Results: 27 studies satisfied the eligibility criteria and were included in the review. The reliability, validity and both reliability and validity were investigated by sixteen, two and nine studies respectively. 17/27 studies were deemed to be of high quality. In total, 15 methods of thoracic kyphosis were evaluated in retrieved studies. All investigated methods showed high (ICC $\geq$ .7) to very high (ICC $\geq$ .9) levels of reliability. The validity of the methods ranged from low to very high.

Conclusion: The strongest levels of evidence for reliability exists in support of the Debrunner kyphometer, Spinal Mouse and Flexicurve index, and for validity supports the arcometer and Flexicurve index. Further reliability and validity studies are required to strengthen the level of evidence for the remaining methods of measurement. This should be addressed by future research.

1. Introduction

Thoracic kyphosis is the sagittal plane curvature between T1 and T12 vertebral bodies (Perriman et al., 2010). Normal kyphosis ranges from 20 to 50° when assessed radiographically (Willlner, 1981). Excessive thoracic kyphosis, defined as a kyphosis >50° (Willner, 1981; Teixeira and Carvalho, 2007), has been previously linked with a range of negative consequences. The postural effects of excessive kyphosis include musculoskeletal complaints such as shoulder pain (Gray and Grimsby, 2004) and cervical pain (Horter, 1978; Callet, 1991; Ayub, 1991) and can affect any age group (Gray and Grimsby, 2004). In osteoporotic samples, excessive thoracic kyphosis can lead to physiological adaptations such as impaired respiratory function (Murray et al., 1993; Di Bari et al., 2004) and can have functional influences such as decreased mobility (Lydick et al., 1997), injurious falls (Kado et al., 2007) and loss of independence (Lydick et al., 1997). The measurement of thoracic kyphosis is therefore an essential aspect to musculoskeletal assessment, helping clinicians to adequately screen for excessive kyphosis, determine baseline data, monitor progress and guide appropriate implementation of treatment strategies (Chaise et al., 2011).

The current gold standard for the quantification of thoracic kyphosis is the lateral radiograph, a method which provides a Cobb angle (Harrison et al., 2001; Briggs et al., 2007). While this is routinely used for the diagnosis and monitoring of conditions such as idiopathic scoliosis and hyperkyphosis (Saad et al., 2012), it has significant limitations. Radiographic methods are generally inconvenient in a clinical setting, involve high costs and expose the patient to high doses of potentially harmful radiation (Korovessis et al., 2001; Kellis et al., 2008). Furthermore, the validity of the Cobb angle has been criticized, particularly in osteoporotic individuals, as it predominantly reflects endplate tilt of vertebrae between selected limits of the curve and fails to represent the full contour of the thoracic spine (Goh et al., 1999; Harrison et al., 2001; Briggs et al., 2007).

* Corresponding author. Tel.: +353 (0)61 234232; fax: +353 (0)61 234251.
E-mail addresses: evabarrett@live.ie, Eva.Barrett@ul.ie (E. Barrett).

1356-689X/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.math.2013.09.003

Alternatively, several non-invasive, skin-surface methods have been adopted for clinical use including the Debrunner kyphometer (Öhlén et al., 1989), the Flexicurve (Milne and Williamson, 1983), the Spinal Mouse (Mannion et al., 2004) as well as technology based methods including rasterstereography (Melvin et al., 2010) and 3D ultrasound (Folsch et al., 2012). Guidelines for selecting measurement tools for use in clinical and research practice recommend that validity and reliability are amongst the essential properties to be considered (Lohr, 2002; Terwee et al., 2007). Validity is an evaluation of whether an instrument measures a construct or variable that it is intended to measure (Carmines and Zeller, 1979; van de Ven-Stevens et al., 2009). For a non-invasive tool to be considered accurate enough to measure thoracic kyphosis in practice and research, it must display adequate criterion validity when compared to the gold standard, i.e. the radiographic Cobb angle. Reliability is defined as the extent to which a measurement is consistent and free from error, when used by the same rater (intra-rater reliability), or when used by different raters (inter-rater reliability) (Portney and Watkins, 2000). In practice, to state that a patient's clinical status has changed since the last measurement, the measured change is required to be larger than the error associated with the measurement (Wright and Feinstein, 1992). Therefore, the reporting of Standard Error of Measurement is an important element of reliability studies as it aids clinical interpretability of results (van de Ven-Stevens et al., 2009).

Since numerous studies on the psychometric properties of these instruments have been published, an evaluation of the literature is required. Therefore, the purpose of this systematic review is to report on the reliability and validity of methods of non-invasive thoracic kyphosis measurement.

2. Methods

2.1. Search strategy

A systematic search was performed on 1st October 2012 by the primary investigator. Searches of the following databases were performed: MEDLINE, AMED, CINAHL, Pubmed, Biomedical Reference Collection: Expanded, SportDiscus, ScienceDirect, Cochrane Library, Web of Science (1960–Oct 2012). The search was conducted using search terms from 3 subject areas: thoracic kyphosis ("thoracic kyphosis", "spinal curvature", "thoracic curvature", kyphosis), psychometric properties (reliability, validity, sensitivity, responsiveness, properties) and physical tests (instrument, tool, test, measure*, inclinometer, flexicurve, kyphometer, radiograph, Cobb). The Boolean Operators "Or" and "And" were used to combine the search terms within and between each of the 3 subject areas respectively. A word from each area was required to be in the Title or Abstract of the study. An additional search of Google Scholar search engine was also performed. These searches were supplemented by hand-searching the reference lists of the final articles found from the above searches.

2.1.1. Eligibility criteria

A meeting between the two reviewers was convened to decide on selection criteria.

2.1.2. Inclusion criteria

- Articles available in full text
- Articles available in English
- A neutral thoracic kyphosis value angle was recorded
- Measurement of validity and/or reliability was the primary aim of the study
- Studies on human participants were included for review. No restrictions were made with regard to populations.

2.1.3. Exclusion criteria

- Full text in English could not be located
- Thoracic kyphosis angle reported in thoracic flexion or extension only
- Radiographic measurement techniques only

Initially, article titles and abstracts were screened by the primary reviewer. Any title and abstract which was not clearly investigating a psychometric property of a thoracic kyphosis measurement method was discarded as being not relevant. In cases of uncertainty about eligibility of a study title/abstract, the full text was explored. When the original search was narrowed down to relevant articles only, a second reviewer independently applied the selection criteria to the chosen articles to ensure all articles were suitable for review. There were no disagreements between reviewers regarding the eligibility of chosen articles.

2.2. Quality assessment

The critical appraisal tool used was a relatively new checklist (Brink and Louw, 2011) which was designed for testing combined reliability and validity studies or validity and reliability on their own. The checklist, which is comprised of 13 items, does not report a quality score. This tool was developed from two existing tools, the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) and the Quality Appraisal of Diagnostic Reliability Studies (QAREL). As some of the included studies assess both reliability and validity of the instrument, this checklist was more convenient than using the QUADAS or QUAREL separately. The criteria are provided as a footnote to Table 2. The studies were considered of high quality if they scored ≥60%, as done previously (van der Wurff et al., 2000; May et al., 2006; 2010; Adhia et al., 2012).

Quality assessment was performed independently by two reviewers on each paper. In the pilot stage, each reviewer independently rated two non-included articles using the checklist, in order to identify any difference in interpretations of the items. This process recorded a kappa score of .92, which was regarded as acceptable to continue. Disagreements were resolved by discussion and all items were clarified.

2.3. Data analysis

Meta-analysis was not attempted due to the heterogeneity of tests, participants and analyses. Also, a subgroup analysis could not be performed due to the limited number of studies evaluating the same thoracic kyphosis measurement technique. Hence a descriptive analysis was conducted and data were synthesized using a level of evidence approach (van Tulder et al., 2003), displayed in Table 1.

The Intraclass Correlation Coefficient (ICC) and Pearson's Correlation Coefficient were interpreted as follows: .00–.29 as very

Table 1
Levels of evidence approach (van Tulder et al., 2003).

Level of evidence	Criteria
Strong	Consistent findings from ≥3 high quality studies
Moderate	Consistent findings from at least 1 high quality and one or more low quality studies
Limited	Consistent findings in ≥1 low quality studies or only 1 study available
Conflicting	Inconsistent evidence in multiple studies irrespective of study quality
No evidence	No studies found