The Reliability and Validity of Upper Quadrant Posture and Two-dimensional Range of Motion Measurement Tools

ABSTRACT: Measuring upper quadrant posture and movement is a challenge to researchers and clinicians. A range of postural measurement tools is commonly used in the clinical setting and in research projects to evaluate postural alignment, but information about the validity and reliability of these tools and thus a selection of the optimal tool for a specific project is often uncertain. This review aims to make recommendations to clinicians and researchers regarding practical, valid and reliable tools to assess upper quadrant posture and range of motion. Electronic databases and key journals were searched. An adapted appraisal tool was utilised to assess the methodology for each of the nine selected articles. Nine eligible articles reporting on the goniometer, flexicurve and inclinometer were included. This review highlights the fact that a range of two-dimensional (2D) posture measurement tools are being used in clinical practice and research. Although the findings for the reliability and validity of the tools included in this review appear to be promising, strong recommendations are limited by the imprecision of the results. Thus, the primary issue hampering the recommendation for the most reliable and valid tool to use in the clinical or research setting is due to the limitations pertaining the analysis of the data, and the interpretation thereof.

KEYWORDS: RELIABILITY, VALIDITY, POSTURE, MEASUREMENT TOOL AND UPPER QUADRANT.
of motion tools. The second objective is to assess the validity of common 2D posture measurement tools.

Primary research reports published in the English language into validity and reliability of upper quadrant 2D posture and cervical range of motion in adults and adolescents without skeletal disease or serious pathology were considered for this review.

4.1 Search Strategy

No date limits were applied to any of the databases. The researcher developed search strategies for Pubmed (since 1950), CINAHL (since 1982), The Cochrane Library (2006 Issue) Science Direct (since 1823) and Embase. The following key words present a summary of the important elements of the search strategy: reliability, validity, photography, digital image, photographic evaluation, photographic analysis, goniometer, inclinometer, electrogoniometer, flexicurve, posture and spine. No MESH headings were used.

In addition, Pearling (searching reference lists) was conducted to identify potential eligible articles from the reference list of eligible articles. A citation search was also done in Pubmed for the authors who have published extensively in the field of posture and cervical range of motion measurement.

The principle researcher and an assistant conducted all the searches and identified eligible articles independently using the defined search strategies for each database. Both the researcher and assistant also independently identified eligible articles. Discrepancies in study selection were discussed till consensus was reached.

4.2 Methodological Appraisal

Nine research articles were identified in the literature search. An adapted Crombie Appraisal Tool (CAT) for review studies was utilised to assess each of the selected articles (Crombie 1996). A review by Katrak et al. (2004) of critical appraisal instruments highlighted the lack of a Gold Standard instrument, and encouraged reviewers to construct instruments that were relevant to their own review purpose (Katrak et al 2004). Questions in the CAT that were inappropriate of this critical appraisal were excluded. The open-ended questions were rephrased in order to allow for dichotomous scoring.

The researcher and assistant appraised the eligible publications independently. Any discrepancies were discussed until consensus was reached and a third party (the study supervisor) was consulted when required. The revised-version of the CAT contained the following questions:

**Figure 1: Database search method and results.**

- Pubmed (n = 83)
- CINAHL (n = 48)
- Embase (n = 232)
- Sciencedirect (n = 28)
- Cochrane (n = 9)
- Author Search (n = 5)
- Reference Search (n = 1)

406 Titles were screened by two independent reviewers

91 Abstracts were retrieved and read by two independent reviewers

32 Full text articles retrieved and read by two independent reviewers

Total n = 9
1. Is the subject of the study clearly stated?
2. Are the aims clearly stated?
3. Is the design appropriate to the stated aims?
4. Was the sample size justified?
5. Are the statistical measures described?
6. Do the numbers add up?
7. Were the basic data adequately described?
8. Is the meaning of the main findings explained?
9. Are factors that might have influenced the observed outcome, discussed?
10. Are important findings overlooked?
11. Is it stated how the results compare with previous reports?
12. Are the implications that the study has for your practice explored?

5. RESULTS

5.1 Selection of Studies and Evidence Level
The findings of the search are presented in Figure 1. Nine eligible articles were included in the systematic review (Harrison et al 2005, Himmann 2003/2004, Lee et al 2003, Malmström et al 2003, Pringle 2003, Tousignant et al 2000, Tousignant et al 2001, Youdas et al 1991, Moffet et al 1989). All nine research studies applied an observational design and are thus ranked as Level 3 on Sackett’s Evidence Hierarchy (Sackett et al 2000).

5.2 Methodological Appraisal

* Sample size calculation
Criterion 4 related to sample size calculation was only fulfilled in one of the studies reviewed (Tousignant et al 2000). Tousignant et al. (2001) compared their sample size of 44 subjects to the sample size of a study conducted by Donner and Eliasziw (1987) who stated that for an 80% power of testing and a 5% of significance, a minimum of 34 subjects is necessary (Donner and Eliasziw 1987).

* Factors that might have influenced the observed outcome
All nine of the reviewed studies obtained a positive answer to this criterion. This criterion was related to the factors that might have influenced the outcome of study. All of the studies considered discussed factors such as an appropriate design of the studies and standardised methodological procedures.

* Description of basic data
Moffet et al. (1989) is the only study that scored a negative response to this criterion as the study did not report on the age of the subjects used (Moffet et al 1989).

5.3 Study Characteristics
In reference to Table 1, the sample size ranged from 26 to 96 and age (M: 47.7, SD: 22.3) ranged from 17 to 88 years old. Two of the studies reviewed used only female subjects while the other seven studies included males and females. None of the studies included adolescents or children.

The earliest study was conducted in 1989 and the most recent in 2005. Fifty-five per cent of the studies were conducted in the United States of America and this may be reflective of research activity or publication bias. None of the published studies included Australian or African populations.

5.4 Study Aims
The primary aim for each of the nine studies is described in Table 2. The three clinical measurement tools included were the goniometer, flexicurve and inclinometer.

Five of the nine studies reported on inter and intra-tester reliability (see Table 2). One of the studies reported inter-tester reliability. The study by Malmström et al. (2003) is the only study reporting on measurement tool reliability (Malmström et al 2003). Concurrent validity was evaluated in three of the reviewed studies while only one study reported on criterion validity of the measurement tool. Harrison et al. (2005) and Lee et al. (2003) Tousignant et al. (2000) determined concurrent validity using X-rays as the gold standard. Harrison et al (2005) included

---

Table 1: Study Characteristics

<table>
<thead>
<tr>
<th>Author</th>
<th>Yr</th>
<th>Sample Size</th>
<th>Age range</th>
<th>Total males</th>
<th>Total females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moffet et al.</td>
<td>1989</td>
<td>26</td>
<td>Not stated</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>Youdas et al.</td>
<td>1991</td>
<td>60</td>
<td>21-48</td>
<td>21</td>
<td>39</td>
</tr>
<tr>
<td>Tousignant et al.</td>
<td>2000</td>
<td>31</td>
<td>18-45</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>Tousignant et al.</td>
<td>2001</td>
<td>44</td>
<td>18-73</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Lee et al.</td>
<td>2003</td>
<td>35</td>
<td>18-35</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Malmström et al.</td>
<td>2003</td>
<td>60</td>
<td>22-58</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>Pringle</td>
<td>2003</td>
<td>27</td>
<td>21-41</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>Hinmann</td>
<td>2003/2004</td>
<td>51</td>
<td>21-51 and 66-88</td>
<td>0</td>
<td>51</td>
</tr>
<tr>
<td>Harrison et al.</td>
<td>2005</td>
<td>96</td>
<td>Mean age:  * Males 17.9  * Females 40.1</td>
<td>36</td>
<td>60</td>
</tr>
<tr>
<td>Study</td>
<td>Measurement Tool</td>
<td>Reliability and Validity</td>
<td>Primary Aim</td>
<td>Static Angle Measured</td>
<td>Study Procedure</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Pringle 2003</td>
<td>Goniometer</td>
<td>Concurrent validity</td>
<td>To compare the tester reliability of the static cervical angle using four different goniometers.</td>
<td>Cervical flexion, extension, side flexion and rotation position</td>
<td>Measurements were done three times with each of the following four devices on the same day: * single hinge inclinometer * single bubble carpenter's inclinometer * dual bubble goniometer * Cybex EDI 320 electrical inclinometer Position of subjects: Standing</td>
</tr>
<tr>
<td>Youdas et al. 1991</td>
<td>Goniometer</td>
<td>Inter and intra-tester reliability</td>
<td>To determine inter and intra-tester reliability measuring static cervical angle.</td>
<td>Cervical flexion and extension position</td>
<td>All subjects were tested thrice in one day by three different testers. Position of subjects: Sitting</td>
</tr>
<tr>
<td>Tousignant et al. 2000</td>
<td>Goniometer</td>
<td>Concurrent validity</td>
<td>To estimate the concurrent validity of the goniometer with x-rays.</td>
<td>Cervical flexion, extension and side flexion position</td>
<td>A lateral cervical X-ray was taken immediately after the measurements were done with the goniometer. Position of subjects: Sitting</td>
</tr>
<tr>
<td>Tousignant et al. 2001</td>
<td>Inclinometer</td>
<td>Inter and intra-tester reliability</td>
<td>To determine the inter- and intra-tester reliability of the inclinometer.</td>
<td>Cervical flexion and extension position</td>
<td>Two measurements were taken by two trained testers. Position of subjects: Standing</td>
</tr>
<tr>
<td>Lee et al. 2003</td>
<td>Inclinometer</td>
<td>Inter and intra-tester reliability Concurrent validity</td>
<td>To establish the inter- and intra-rater reliability and validity of the inclinometer using thoracic X-rays.</td>
<td>Thoracic flexion, extension and side flexed position</td>
<td>Two raters took single inclinometry measurements on two different days. The same angle was captured on x-rays and compared. Position of subjects: Standing</td>
</tr>
<tr>
<td>Malmström et al. 2003</td>
<td>Inclinometer</td>
<td>Measurement tool reliability Inter and intra-tester reliability Concurrent validity</td>
<td>To estimate the measurement tool reliability, concurrent validity, inter-tester and, intra-tester reliability of the inclinometer. Validity assessment using ultrasound.</td>
<td>Cervical flexion and extension position</td>
<td>Recordings were made with the following inclinometers: * Zebris three dimensional ultra-sound motion device * Myrin gravity-reference goniometer, simultaneously Position of subjects: Standing</td>
</tr>
<tr>
<td>Moffet et al. 1989</td>
<td>Inclinometer</td>
<td>Inter and intra-tester reliability</td>
<td>To determine the inter- and intra-tester reliability of the inclinometer with static cervical angles.</td>
<td>Cervical flexion, extension, right and left side bending and rotation position</td>
<td>Neck angles were measured three times in one hour by the same observer. Neck angles were measured by two observers at the same time. Position of subjects: Standing</td>
</tr>
<tr>
<td>Hinmann 2003/2004</td>
<td>Flexicurve</td>
<td>Inter-tester reliability</td>
<td>To establish the inter-rater reliability of the flexicurve.</td>
<td>Thoracic kyphosis and lumbar lordosis.</td>
<td>Three graduate students measured cervical lordosis and thoracic kyphosis in normal standing posture and then in an erect posture. Position of subjects: Standing</td>
</tr>
<tr>
<td>Harrison et al. 2005</td>
<td>Flexicurve</td>
<td>Concurrent validity</td>
<td>To validate the flexicurve contour measurements of the cervical spine lordosis with cervical X-rays.</td>
<td>Cervical lordosis</td>
<td>The flexicurve skin contour and neutral lateral x-rays were digitized and compared. Position of subjects: Standing</td>
</tr>
</tbody>
</table>
96 adult subjects and Lee et al. (2003) included 20 adult subjects (Harrison et al. 2005, Lee et al. 2003). The inclusion of only adults may be due to radiation exposure. The relatively small sample by Lee et al. (2003) may be due to the economic costs related to radiation exposure.

One of the reviewed studies reported on subject variability (Malmström et al. 2003). This is viewed to be a shortcoming as it is an important element in estimating the standard error of measurement. This element should thus be considered in future studies evaluating reliability of 2D posture measurement tools, particularly considering the individual variability in posture (Christensen 1999).

5.5 Methodological Procedures


A variety of angles were measured in the studies reviewed. This included cervical flexion, extension, side flexion, left and right rotation static angles. Thoracic measurements included thoracic flexion, extension, side-flexion and rotation static angles. The studies were primarily conducted to evaluate elements of reliability and validity by measuring the static position after subjects were instructed to place their necks or thoracic spines in a specific position such as flexion or extension.

The only study which reported on the postural alignment of body segments is the study by Hinmann 2003/4. The researchers measured the position of lumbar lordosis and lumbar kyphosis.

5.6 Study Findings

5.6.1 Reliability Results

The goniometer was found to be reliable in two studies (see Table 3) (Pringle 2003, Youdas 1991). The Inter-class Correlation Coefficients (ICC) values for the plastic hinge goniometer ranged between 0.89 for flexion and 0.97 for cervical side-flexion. The single goniometer scored ICC values of between 0.79 for cervical side flexion and 0.92 for flexion-extension combined.

The dual bubble goniometer scored a minimum ICC value of 0.86 for the flexion-extension combination and a maximum of 0.94 for the side-flexed position. The cybex electric goniometer only assessed the static ROM for flexion-extension combination (ICC=0.89) and the side bending position was (ICC=0.75). Youdas et al. (1991) found the goniometer to be moderately reliable with ICC
values ranging between 0.78 and 0.90 for the intra-tester reliability and 0.54 and 0.79 for the inter-tester reliability (see Table 3) (Youdas et al 1991).

The reliability of the inclinometer was determined by four different studies. The first study found the inclinometer to have good inter-tester reliability (Malmström et al 2003). The second and third studies found it to have moderate reliability (see Table 3) (Tousignant et al 2000, Moffet et al 1989) and the fourth (Lee et al 2003) (see Table 3) found it to be completely unreliable.

Himmann 2003/4 determined the inter-tester reliability of the flexicurve (Himmann 2003/2004). For the kyphosis in relaxed posture, an ICC value of 0.94 was obtained for the erect posture, 0.93. For the lordosis in relaxed posture, an ICC of 0.60 was obtained and for the erect posture, 0.73.

5.6.2 Validity Results
Table 4 demonstrates the validity of the inclinometer, goniometer and flexicurve according to the study findings. The inclinometer was found to be valid in one study (Malmström et al 2003) and invalid in another (Lee et al 2003). The goniometer demonstrated excellent validity for what cervical movements (Tousignant et al 2001) whilst the flexicurve was found to be invalid for cervical lordosis measurement (Harrison et al 2005).

6. DISCUSSION
This is the first attempt to review the reliability and validity of the goniometer, inclinometer and electrogoniometer when assessing the postural alignment of body segments of the upper quadrant.

This review demonstrated that only one published study provides information on the validity of postural alignment. Harrison et al (2005) assessed concurrent validity of the flexicurve comparing it to X-rays when measuring cervical lordosis in the standing position (Harrison et al 2005). There is thus a dearth of literature reporting on the validity and reliability of “natural” postural alignment, despite the fact that posture is a predictor of musculoskeletal dysfunctions during sedentary activities such as sitting while using a computer (Vieira and Kumar 2004, National Institute for Occupational Safety and Health). Considering that posture is frequently assessed and rehabilitated by physiotherapists, further research in this field is required.

Seven of the nine studies performed the measurements on subjects who were standing (Harrison et al 2005, Himmann 2003/2004, Lee et al 2003, Malmström et al 2003, Pringle 2003, Tousignant et al 2000, Moffet et al 1989), and only two of the reviewed studies tested subjects in a sitting posture (Tousignant et al 2001, Youdas et al 1991). From this data, there appears to be an inadequate understanding of the reliability and validity of sitting posture measurement. This is worrisome to physiotherapists who would like to use these measures as most of our clients spend significant time in a seated position, either at school or at work, in a car or in front of the television (Granjean and Hunting 1977). Thus, it is proposed that further studies should be carried out where posture and movements are assessed, in all functional postures, to improve the clinical utility of the findings.

The goniometer was the most studied instrument and the reliability results appear to be promising. Unfortunately, the methodological limitation in terms of sample size and statistical analysis does not allow us to strongly recommend the goniometer as a reliable tool. Firstly, although most of the studies appropriately applied the ICC, if the sample size is too small, the confidence intervals would be very wide (Jordan et al 2000). All, except one of the studies, calculated the sample size and only five of the reliability studies had a sample size of more than 30 subjects. It is a general guide that a sample size of at least 32 is required based on three repetitions or observers to give a power of more than 80%. Therefore, if no sample size calculation is provided it is crucial that confidence intervals are provided to allow for confidence in the study findings. In this review, none of the studies provided confidence intervals and therefore the degree of confidence in the findings is questionable.

Concurrent validity was evaluated in three of the reviewed studies while one study reported on criterion validity of the measurement tool. Harrison et al (2005) and Lee et al. (2003) determined concurrent validity using X-rays as the gold standard (Harrison et al 2005, Lee et al 2003). Normal X-rays pose a health threat to children and adolescents (Wall et al 2006). Children are considerably more sensitive to the carcinogenic effects of ionizing radiation than adults, and children have a longer life expectancy in which to express risk.

The validity studies demonstrated similar shortcomings with respect to statistical analyses. None of the validity studies reported confidence intervals and therefore, it is not possible to ascertain whether these tools will be valid in a given population. Furthermore, all studies only used ICC’s which has limitations in providing insight to clinicians regarding the accuracy of the tool compared to the gold standard. ICC values are difficult to conceptualise as it is not in the unit of measurement and therefore does not provide clinicians with information on which they can base the selection of a specific tool. It is thus recommended that Bland Altman plots should also be calculated as this gives the degree of accuracy in the specific measurement unit, such as degrees, and thereby allows much better understanding of the validity of the tool.

7. CONCLUSION
This review highlights that a range of 2D posture measurement tools are being used in clinical practice and research. Currently, the majority of the reviewed publications report on the reliability and validity of the goniometer. Although the findings for the reliability and validity of the tools included in this review appear to be promising, strong recommendations are limited by imprecision of the results. Thus, the primary issue hampering the recommendation for the most reliable and valid tool to use clinically or in a research setting is limitations pertaining to the data analysis and interpretation.

ACKNOWLEDGEMENTS
The principle author would like to thank Prof. Quinette Louw for guidance and support.
REFERENCES:


