The reliability of video fluoroscopy, ultrasound imaging, magnetic resonance imaging and radiography for measurements of lumbar spine segmental range of motion in-vivo: A review

Eleanor Shalini Daniel

Faculty of Health and Social Sciences, Bournemouth University, Bournemouth, UK

Raymond Y W Lee

Faculty of Technology, University of Portsmouth, Portsmouth, UK

Jonathan Mark Williams

Faculty of Health and Social Sciences, Bournemouth University, Bournemouth, UK

Corresponding Author: Eleanor Daniel, Faculty of Health and Social Sciences, Bournemouth University, Bournemouth Gateway Building, St Pauls Lane, Bournemouth, BH8 8GP. E-mail: edaniel@bournemouth.ac.uk

The reliability of video fluoroscopy, ultrasound imaging, magnetic resonance imaging and radiography for measurements of lumbar spine segmental range of motion in-vivo: A review

Abstract

Background: Lower back pain (LBP) is a principal cause of disability worldwide and is associated with a variety of spinal conditions. Individuals presenting with LBP may display changes in spinal motion. Despite this, the ability to measure lumbar segmental range of motion (ROM) non-invasively remains a challenge.

Objectives: To review the reliability of four non-invasive modalities: Video Fluoroscopy (VF), Ultrasound imaging (US), Magnetic Resonance Imaging (MRI) and Radiography used for measuring segmental ROM in the lumbar spine in-vivo.

Methods: The methodological quality of seventeen eligible studies, identified through a systematic literature search, were appraised.

Results: The intra-rater reliability for VF is excellent in recumbent and upright positions but errors are larger for intra-rater repeated movements and inter-rater reliability shows larger variation. Excellent results for intra- and inter-rater reliability are seen in US studies and there is good reliability within- and between-day. There is a large degree of heterogeneity in MRI and radiography methodologies but reliable results are seen.

Conclusions: Excellent reliability is seen across all modalities. However, VF and radiography are limited by radiation exposure and MRI is expensive. US offers a non-invasive, risk free method but further research must determine whether it yields truly consistent measurements.

Keywords: kinematics, back, spine, measurement, reliability.

1. Introduction

Lower back pain (LBP) is the principal cause of disability worldwide and the sixth leading contributor to overall disease burden [1]. LBP affects approximately 540 million people globally at any one time [2]. International studies have reported LBP point prevalence rates between 12 and 35% and lifetime prevalence rates ranging from 49 to 80% [3]. As a result, LBP is one of the most common reasons for an individual to seek medical attention [4]. In the United Kingdom alone, the estimated direct cost of healthcare for LBP exceeds £1 billion per year [5].

Despite this substantial economic burden, the pathophysiology of LBP is poorly understood [6]. However, evidence suggests individuals commonly display differences in movement behaviour [7–9], some of which are believed to reflect changes in segmental spinal motion [10,11]. In support of this, Haxby-Abbott et al. (2006) [12], demonstrated that LBP was associated with a reduction in segmental sagittal range of motion (ROM). In comparison, Kulig et al. (2007) [10], found that LBP was associated with an increase in segmental sagittal ROM.

Therapeutic models of LBP assessment and treatment across a range of professions are firmly embedded in this notion of change in segmental ROM. In addition, segmental ROM assessment is also critical for enhancing the understanding of existing spinal diseases, aiding spinal diagnoses and evaluating contemporary treatment or surgical intervention. For these reasons, the measurement of lumbar spinal ROM is clinically important [13], yet the ability to measure an individual's segmental ROM non-invasively remains a challenge [14].

Kinematics of the lumbar spine have been studied using a range of techniques including implantable bone pins [15,16] and implanted ball bearings [17]. However, due to the invasive nature of these methods, they are unlikely to become routine clinical

practice. Non-invasive methods including radiography [18], video fluoroscopy [19,20], magnetic resonance imaging [21,22] and ultrasound [23] are alternate methods reported in the literature. However, to date, no contemporary synthesis of the literature exploring these non-invasive methods to assess segmental ROM has been completed. Understanding these current methods will provide insight into, and future direction for, the tools required for exploration of long held segmental ROM notions and a step change in the use of imaging for spinal pathologies.

The purpose of this study was to review the reliability of four current non-invasive modalities (Video Fluoroscopy (VF), Ultrasound (US) imaging, Magnetic Resonance Imaging (MRI) and Radiography) used for measuring segmental ROM in the lumbar spine in-vivo, through systematic examination of the literature. This study will form a definitive reference resource for clinical research into segmental ROM measurement aiding clinical researchers in selecting the most appropriate measurement methods for their application.

2. Materials and Methods

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [24].

2.1 Search Strategy

In January 2021, a systematic literature search of electronic databases including: CINAHL complete, Academic Search Ultimate, MEDLINE Complete, ScienceDirect, Complementary Index, PsycINFO and Supplemental Index was conducted using key terms and Boolean logic for each modality, as listed in Table 1. Each search was limited to peer-reviewed articles, published in the English language. Table 1 shows the number

of articles yielded for each modality after exact duplicates were removed.

Articles were initially screened by title, abstract, and where necessary full text, against inclusion and exclusion criteria (as listed in Table 2) by the first author; with any uncertainty resolved by consensus (ED, JW). All studies deemed appropriate for this review were also checked and confirmed by an additional author (JW). A detailed flow chart of the search can be seen in Figure 1.

2.2 Inclusion and exclusion criteria

Studies needed to investigate segmental ROM of the lumbar spine in-vivo (human participants) using VF, US, MRI or Radiography. Consideration of the modality's psychometric properties was also required by the articles. For the purpose of this review this means studies had to explore characteristics of reliability and validity, such as repeated measures reliability and estimates of error. See Table 2 for detailed inclusion and exclusion criteria and Table 3 for reasons for article rejection.

2.3 Quality assessment

Critical appraisal of the methodological quality of each article was completed by the first author using an assessment tool for observational cohort and cross-sectional studies, taken from the National Heart, Lung and Blood Institute [25]. The appraisal criteria consisted of 14 questions that could be answered yes, no, cannot determine, not applicable or not reported. Then, an overall quality rating was given based on these answers. The results can be seen in Table 4. This tool was used because its design draws focus to the key concepts of a study; facilitating evaluation of its internal validity [25].

3. Results

A total of 17 studies were eligible for this review [23,26–41]. Table 5 summarises the data extracted and Table 6 summarises the findings.

Six studies used VF to measure ROM [28,33–35,38,40], five used radiography [26,27,29,30,32], four articles used US imaging [23,36,37,39], and two citations investigated MRI [31,41]. However, Chleboun et al. [23] also included MRI results as a gold standard comparator for US.

Overall, 600 participants were included in this review; of which at least 289 were male and 243 were female. Two studies [26,41] did not report the breakdown of male to female participants. 250 participants were symptomatic whilst 350 were classed as healthy or asymptomatic.

Most studies involved only healthy participants [23,26,28,36–40], whereas some had a mixture of symptomatic and asymptomatic individuals [33–35,41] and others studied specific populations [27,29–32]. These included participants with LBP [27,31], spondylolisthesis [32], monosegmental degenerative disc disease [29] and monosegmental total disc replacement [30].

Articles measured segmental ROM during flexion and/ or extension [23,27,29–36,38]. Others investigated flexion, extension and side flexion [28,40], two studies [37,39] quantified motion of the lumbar spine from three static positions; whilst one study looked at neutral positioning and lateral bending motion [26].

The psychometric properties of each modality analysed varied between reliability [23,26–30,32–35,38–41] or a combination of reliability and validity [31,36,37]. All but five studies [26–30] used intra-class correlation coefficient (ICC) as a metric of reliability. Additional outcomes studied amongst the articles were standard error of measurement (SEM) [26,31,32,34–36,38], co-efficient of variation (CoV)

[23,41], pearson correlation coefficient (PCC) [29,30], kappa [26], root mean square error (RMSE) [28] and minimal detectable change (MDC) [36,37,39,40].

All studies had an overall quality rating of fair or good based on the 14-point appraisal checklist [25] but demonstrated similar methodological flaws and thus, shared common threats to validity.

3.1 Methodological Analysis

Only three studies [34,36,38] justified their sample size or provided a description of study power. This methodological element is important to ensure an adequate number of participants are studied to yield valid estimates of reliability. As sample size varies considerably across the studies, it is likely that the power also varies significantly and this should be considered when extrapolating the findings.

All studies, excluding five [29,30,38,40,41] took ROM measurements from images at only one stage during the study period, thus exploring within-day repeated measures reliability. Whilst this is likely to result in more consistent movement patterns; conclusions regarding reliability of between-day repeated movements are not possible.

Additionally, aside from two studies [34,40], key potential confounding variables were not reported. Confounding factors are characteristics which may influence the dependant variable and thus, alter the findings of a study. For example, US imaging can be more difficult in individuals with a high body mass index (BMI) [42] and likewise, this category of participants may require a stronger radiation dose for VF [43] and radiographs [44]. Similarly, the quality of MRI images can be affected by permanent cosmetics, including tattoos [45]. In the absence of the consideration of confounding factors, it is difficult to determine if their presence or absence affected the results.

3.2 Reliability

3.2.1 Video Fluoroscopy

3.2.1.1 Segmental ROM values

Segmental ROM of flexion across the studies were similar, ranging from 4.05º to 7.10º for lying [34,35,40] and 9º to 14º for standing [40]. Extension has been less frequently studied with segmental ROM values of 4.11º to 5.31º [34], 2.00º [35], 5.33º for lying and 2.01º in standing [40]. Landel et al. [31] and Sui et al. [33] did not report individual segmental ROM values.

3.2.1.2 Intra-rater reliability of segmental ROM measurement

In VF, automated tracking algorithms are commonplace; where bony boundaries are automatically tracked by a computer from which ROM calculations are made [46]. In most cases, an operator is required to manually mark the first image, or few images, from which the tracking algorithm commences [47]. This manual identification is known to be an important source of error both between individuals and within individuals [48]. To this end, a body of work has concentrated on quantifying the variability this manual marking of images affords [28,46,48]. The methodology involves participants completing one movement in the fluoroscope, from which multiple mark ups and analysis are completed. This is either repeated by the same individual or between individuals.

Using a mixture of individuals with pathology and those asymptomatic, Yeager et al. [35] demonstrated excellent reliability (ICC 0.98, CI95% 0.98-0.99, SEM 0.10º, SEM% = 2% for flexion, 5% for extension) for the same investigator repeatedly

marking-up and processing the same VF sequences. These included sagittal plane motions only and were a mixture of upright and recumbent movements. Similar findings were reported by Mellor et al. [34] for lying motion, where excellent reliability was established for sagittal plane motions (ICC 0.92-1.0, CI95% 0.72-1.0, SEM 0.10º to 0.35º, SEM% = 3% flexion, 8% extension). In addition, similar findings have been reported for recumbent side bending movements where ICCs ranged from 0.99-1.0, CI95% 0.95-1.0, SEM 0.08º to 0.17º, SEM% = 3% lateral bending [34].

These results demonstrate that if the same individual marks up and processes the images; VF can be used to reliably measure lumbar sagittal ROM in recumbent and upright as well as, recumbent side-bending with a small SEM.

3.2.1.3 Inter-rater reliability of segmental ROM measurement

Investigation of the inter-rater reliability of processing the same images show sagittal ICC values remain good-to-excellent but are slightly lower for extension (ICC 0.74-0.99, CI95% 0.23-0.99) [34], Yeager et al. [35] (ICC 0.96, CI95% 0.95-0.97). It should be noted that the confidence interval for extension was large; with lower estimates suggesting poor reliability. In addition, the SEM values were also higher at 0.22º (or 5% flexion and 11% for extension) [35], and 0.17º to 0.31º for flexion (or 7%) and 0.27º to 0.77º for extension (or 19%) [34]. It is not clear why Yeager et al. [35] have much lower SEMs compared with Mellor et al. [34] but it is apparent that Mellor et al. [34] contained outliers for extension which could have affected results. Furthermore the presented percentage SEM of 19%, reported at L2/3 by Mellor et al. [34], is the largest expected. It is possible that this was at the upper edge of their fluoroscope and therefore, was affected by image quality. Comparatively, Yeager et al. [35] investigated L1/2 as their upper segment; suggesting a wider field of view. To this end, the L4/5 segment

assessed by Mellor et al. [34] affords much better reliability for extension measurement (ICC 0.99, CI95% 0.96-1.0, SEM 0.27º, SEM% = 5%).

Altogether, these results indicate that larger variation is seen when different individuals process the same VF motion sequences, even though automated algorithms are used. Nevertheless, the ICCs remain good-to-excellent. Furthermore, although some larger errors are noted for extension; errors were small, especially for flexion.

3.2.1.4 Repeated Movements

The measurement of repeated movements is not common in VF research, presumably due to repeated participant exposure to radiation. However, establishing this enables more than just the error in marking up of VF images to be explored. Humans have a natural variance in movement [49,50], and this variance needs quantifying prior to any methods being employed for repeated measures in clinical studies. To date, only one study has investigated this. Breen et al. [28] conducted a baseline measurement and follow-up measurement approximately 30 minutes later. Unfortunately, due to some technical issues, repeated measures reliability was only reported for side bending; with RMS errors of 2.75º to 2.91º [28]. Raw data ranges were not reported but using those from Mellor et al. [34], who had a similar methodology; this represents around 52% error.

As a result, even with the same individual marking up images, this suggests that errors are quite large when exploring repeated measurements with VF.

3.2.1.5 Between-day reliability

To explore between-day reliability some studies have taken a VF sequence, processed it and then reprocessed it sometime later to explore between-day intra-rater reliability

[38]. Excellent reliability for all vertebral levels was established with SEM values between 0.23º and 0.54º [38]. However, this just represents errors associated with processing, rather than the additional biological variation of repeated movements.

This variation has been recently studied in 55 participants and over 200 motion segments, both in lying and standing, without pain or known pathology [40]. ICC values suggest excellent reliability (0.80 for lying flexion and extension, 0.82-0.91 for standing flexion and extension) and small confidence intervals (lowest ICC 95%CI = 0.74) [40]. Rather than reporting SEM, the authors chose to present the MDC at the 95% confidence level (MDC95). The MDC95 values are high suggesting significant variance in the repeated measurements. For example, the MDC95 value for flexion in lying was 4.66º [40]. This means that with 95% confidence, a change greater than 4.66º represents true change beyond normal variation expected with repeated testing. The total range measured was 5.14º [40] thus, a change of 4.66º on 5.14º indicates 91% change is needed before there is confidence that this represents real change. As percentages, the magnitude of MDC95 was 91% for flexion in lying, 97% for extension in lying, 100% for flexion in standing and 176% for extension in standing [40]. Therefore, a change from 9.1º average flexion in standing to over 18º would be required to provide 95% confidence that is was true change beyond natural variation. Previous studies would suggest this may not be physiologically possible or at least, puts the segmental ROM in the top 2% of normal [34]. Similar findings were observed for side bending with good ICC values but high MDC95 values (60-69% lying, 97-98% standing) [40].

In summary, it is clear that within-day reliability of marking up and processing VF sequences is excellent for both intra- and inter-rater. However, the intra-rater reliability of measuring repeated movements within-day demonstrates larger errors, and these are even greater when investigating between-day reliability. Therefore, if using

VF to investigate interventional changes across days, large change in segmental ROM are needed to be sure these are greater than natural variability. This suggests low sensitivity to change of measuring repeated movements with VF.

3.2.2 Ultrasound

3.2.2.1 Segmental ROM values

In order to quantify segmental ROM using US, many studies [23,37,39] opted to visualize and then measure the linear distance between two adjacent spinous processes. Therefore, reporting of segmental ROM was commonly as a linear distance measurement in the units of millimetres (mm). Three studies investigated a 'neutral', flexed and extended position in either prone [37,39] or supine [23]; whilst the other study investigated 'neutral' in standing and forward bending motion [36].

Values for spinous process separation in flexion ranged from 25.6mm to 32.3mm [23] and 29.2mm to 30.1mm [37]. Distance measures for extension ranged from 21.5mm to 26.9mm [23] and were reported only in this study. Actual flexion ROM, taken from neutral, ranged from 3.0mm to 4.4mm and were only reported in one study [23]. Segmental ROM was reported in degrees for Cuesta-Vargas [36] using an image rotation method; yielding values of 15.4º to 16.3º for segmental ROM during flexion.

3.2.2.2 Intra-rater reliability of segmental ROM measurement

Intra-rater reliability estimates were reported as excellent by Chleboun et al. [23] (ICC 0.94, CI95% 0.85-0.97), Tozawa et al. [39] (ICC CI95% 0.963-0.999) and good-to-excellent (ICC CI95% 0.79-1.0, or with one examiner removed CI95% = 0.92-1.0) by

Tozawa et al. [37]. Small coefficient of variances have been reported (1.8%) [23], along with moderate MDC95 values of 0.29mm (around 10%) [37]. However, these could be as large as 1.8mm (around 60%) [39] based on segmental ROM of 3.0mm.

Both Tozawa et al. [37] and Chleboun et al. [23] positioned the participant in one position and collected all three measurements in that same position prior to then moving onto the new position, henceforth eliminating the biological variation of repeated movement measurements. Nevertheless, this method doesn't replicate the type of method required to determine the repeated measures reliability that is more normal in biomechanical studies. This includes the biological variation of the human completing repeated movements.

3.2.2.3 Inter-rater reliability of segmental ROM measurement

Inter-rater reliability was explored in two studies [37,39] with good-to-excellent reliability reported by Tozawa et al. [37] depending on the measurement method (ICC 0.914, CI95% 0.80-0.97; ICC 0.725, CI95% 0.55-0.87) and excellent reliability seen in their follow up study (ICC 0.969, CI95% 0.90-1.00) [39].

3.2.2.4 Repeated movements

Only one study investigated repeated movements (flexion) measured with US [36]. Excellent estimates of reliability were reported for both within-day (CI95% = 0.995-0.999) and between-day (CI95% = 0.996-0.999) [36]. Moreover, MDC95 estimates were made from the SEM reported in the article (SEM = 0.54º, MDC95 = 1.5º or 10%) [36], indicating change greater than 10% would represent true change.

Overall, these US results show that if the same individual captures repeated images without altering the participant's position; excellent intra-rater reliability should

be expected. This expectation is further extended to between individuals. In addition, MDC95 values could be up to 60%, but these have not been established for between individuals. Consequently, this is an important consideration when designing test-retest studies. The values of MDC95 provide estimates as to the sensitivity of change, which is important when designing future experiments. Lastly, repeated movements have been less well investigated but estimates from a single study show promising reliability within- and between-day.

3.2.3 MRI

3.2.3.1 Segmental ROM values

The studies included in this review focussing on MRI often had primary aims not aligned to proving the utility of MRI for segmental ROM testing. Some used it as a gold standard comparator [23], others for validity of manual therapy [31]. Only Mahato et al. [41] focused on segmental ROM.

The distance between spinous processes were reported as a surrogate of flexion and extension with values ranging from 24.6mm to 35.6mm for flexion, 19.9mm to 29.4mm for extension and segmental ROM estimates, from neutral of 1.8 to 4.9mm for flexion and 0.9 to 4.3mm for extension [23]. Actual segmental ROM values for right side bending were reported between 8.5º and 17.3º depending on the segment [41].

3.2.3.2 Reliability

Regarding reliability, a synthesis of the studies is difficult due to a large degree of heterogeneity evident in the methodology.

Chleboun et al. [23] utilised supine positioning with wedge placement to induce extension and flexion and three measures were taken without moving from each position. This method is unlikely to achieve full ROM and it also removes all biological variation due to repeated movement. As a result, reliability estimates were excellent (ICC = 0.98, CI95% = 0.95-0.99; CoV = 1.6%) [23]. Landel et al. [31] completed a prone MRI during manual palpation and 'accessory spinal mobility assessment'. They quantified the curvature change during 'posterio-anterior' pressure. However, they did not report any actual values of curvature. Reliability estimates for intervertebral curvature in prone for five participants across two visits were excellent (ICC 0.95-0.99; SEM 0.40º to 0.66º) [31]. Mahato et al. [41] completed MRI during right side bending between days. ICC estimates for between-day reliability of segmental ROM (side bending) were excellent 0.93-0.94 and CoV values at 14-15% [41].

In summary, regardless of the methods employed, it appears that MRI for segmental ROM measurements is highly reliable in both the sagittal and frontal plane for end of range static positions. Despite this, the coefficient of variation seems to depend on the movement being measured and the method of analysis. Similarly, the effect of different assessors and of true repeated movements is not clear.

3.2.4 Radiography

3.2.4.1 Segmental ROM values

Since the aim of this review was to investigate reliability, the search of radiography papers was limited to those investigating this psychometric outcome. As a result, the citations included in this review are not inclusive of the exhaustive list of radiography studies that report segmental ROM values. Readers interested in this area are directed to

papers such as Yukawa et al. [51] and Galbusera et al. [52].

Measurements of lumbar segmental ROM from radiographs varied between the included studies. Three studies reported at least one plane of lumbar segmental rotation including side bending and rotation [26] and flexion-extension [27,32]. Individual segmental ROM values were not reported in three studies [26,29,30].

Using similar conceptual methods, segmental ROM was quantified from flexion-extension radiographs in two studies by reporting the angle change between adjacent vertebral endplates [27,32]. Pearson et al. [32] found an average change in intervertebral rotation of 5.1º and 5.7º for the digitised manual technique (DMT) and computer-assisted quantitative motion analysis (QMA) method respectively. However, it is not known whether these results are in relation to weight bearing or recumbent postures as they did not detail the positioning of participants during imaging.

Maigne et al. [27] also reported segmental ROM values but in sitting and standing positions of participants with chronic LBP. Some had pain that occurred immediately on sitting down which was relieved on standing up (patient group) and participants who did not have these symptoms were matched to the patient group based on age and gender (control group) [27]. Angular motion (AM) for positional change from extension to flexion was 13.9º±4.5º (patient group) and 7.5º±4.3º (control group) [27]. Similar values were seen for positional change from extension to sitting (AM = 10.0º±4.5º (patient group); 6.2º±4.0º (control group)) [27]. It is not immediately clear why such large ROM was observed in the patient group. However, the influence of LBP on the motion in this group could offer explanation as well as, the sample being largely female and the methods specific focus on achieving maximal ROM.

3.2.4.2 Intra-rater reliability of segmental ROM measurement

In radiography research, reliability analysis usually involves one or several raters measuring segmental ROM from the same radiographs on one or multiple occasions. However, due to variability in methodology and presented reliability statistics, synthesis of the studies included is difficult.

Using two raters and two measurement methods, Cakir et al. [29] and Cakir et al. [30] investigated the intra-rater reliability of measurements from standing flexion-extension radiographs, with measurements taken from the same images on two separate occasions. Intra-rater reliability estimates for segmental ROM were reported as strong for measurements made by the same rater using the same method (PCC = 0.782-0.916) with small mean differences between the two measurements (-0.17º to 0.04º) [29]. However, the 95% confidence intervals for these differences ranged from ±4.0º to ±6.8º) [29] suggesting that despite a small mean, there was a large range of differences between two measurements.

Similar outcomes were observed for their follow up study where the method was adapted to measure the intervertebral segment which had received a total disc replacement [30]. Strong intra-rater reliability estimates (PCC = 0.903-0.962) with small mean differences (-0.08º to 0.08º) were reported but there were wide confidence intervals between these two measurements (±2.0º to ±3.3º) [30].

In Pearson et al. [32] study, 30 flexion-extension radiographs were measured twice by six raters, over a four week period, using either the DMT or QMA method. Intervertebral rotation intra-rater reliability ICCs were higher for the QMA method (ICC = 0.997) with small SEM (0.5º) compared to the DMT (ICC = 0.870, SEM = 2.5º) [32].

For end-plate angle in extension, flexion and sitting, Maigne et al. [27] analysed the intra-rater reliability of one rater extensively by opting to investigate if there was a

difference between repeated measurements. They determined no significant difference between repeated measurement of the same images, reporting that the mean difference between two measurements was ≤0.31º [27]. However, the 95% confidence interval for the limits of agreement between the measures was at best -3.0º to 2.4º suggesting significant variability between repeated measures [27].

3.2.4.3 Inter-rater reliability of segmental ROM measurement

Inter-rater reliability estimates for segmental ROM of flexion-extension radiographs amongst two raters were reported as strong for measurements between raters using the same method (PCC = 0.738-0.929), with a small mean difference (-0.82º to -0.07º) [29]. However, as observed before, the range of difference between two raters was large; yielding a wide 95%CI (-7.4º to 5.8º) [29]. Similar findings were observed in their follow up study with a strong correlation between raters (PCC = 0.886-0.950) and small mean difference (-0.31º to 0.04º) but large confidence intervals (-3.0º to 3.1º) [30].

Inter-rater reliability of flexion-extension radiographs was further studied by Maigne et al. [27] and Pearson et al. [32]. Estimates provided by Maigne et al. [27] demonstrated mean differences between two raters measurements in extension, flexion and sitting was -0.38º to -1.05º. However, the 95%CI between raters was -3.1º to 4.8º for end-plate angle, suggesting wide variability in the differences [27]. Pearson et al. [32] found inter-rater reliability estimates that were excellent for measurements made with the QMA method (ICC = 0.976) compared to the DMT that yielded moderate results (ICC = 0.693).

Haas et al. [26] investigated tilt into side bending and rotation in standing and lateral bending positions using three examiners reporting a range of Kappa reliability estimates. For side bending, agreement between raters was reported as weak-to-

moderate in neutral (Kappa = 0.17-0.56) for L1-L5 [26]. This was also true for net segmental tilt in left lateral bending (LLB) (Kappa = -0.03-0.50) and right lateral bending (RLB) (Kappa = 0.00-0.27) excluding the measurement at L3 for LLB that showed excellent reliability (Kappa = 1.00) [26]. However, the inter-rater reliability estimates were better overall for rotation results which yielded moderate-to-good results in neutral radiographs (Kappa = 0.55-0.68) and weak-to-good results in LLB and RLB (Kappa 0.38-0.68) [26]. Interestingly, reliability estimates at L5 were low across all three raters in neutral, LLB and RLB for tilt (Kappa = 0.16-0.19) and rotation (Kappa = 0.38-0.57) [26].

The SEM was also reported by Haas et al. [26]. They found the mean absolute discrepancy was <2º for tilt and <3º for rotation of neutral radiographs at all lumbar levels [26]. This was less than half of the expected measurement error which was also true for net tilt (1.2º to 3.2º) and rotation (2.0º to 3.7º) in left and right lateral bending [26]. However, even though all measurement errors were reported as low, data were only presented from one rater pair [26].

Overall, results for radiography indicate that there is high intra-rater reliability between measurements made using the same method, and differing methods, in flexion-extension. This also appears true for inter-rater reliability in flexion-extension as well as, lateral bending radiographs. However, variability in the results suggest reliability could be affected by the selected method for measuring ROM from the radiographs. Moreover, the magnitude of the variability across 2 measurements of the same image should be considered when assessing the expected ROM alteration from interventions such as surgery.

4. Discussion

This review set out to provide a contemporary synthesis of the reliability of four current non-invasive modalities used for measuring segmental ROM in the lumbar spine in-vivo. Detailed understanding of current methods is important for researchers as it enables recognition of what systems are available and there associated strengths and weaknesses. This facilitates informed judgements pertaining to the use of such methods, including determining whether or not a method is reliable for its planned application. In addition, this work also served as a valuable reference resource to aid clinicians in the interpretation of clinical findings ensuring, for example, that changes reported in clinical trials are beyond those expected due to natural variability. Understanding these current methods will provide insight into, and future direction for, the tools required for exploration of long held segmental ROM notions and a step change in the use of imaging for spinal pathologies.

4.1 Modality evaluation

4.1.1 VF

VF provides a cost-effective, non-invasive [53] method for segmental ROM assessment that can provide dynamic or static quantification of ROM and is often completed in a weight-bearing position. However, there is a tricky trade-off between radiation dose and image quality [54]. Since low radiation dosage is used [55]; the contrast between the vertebrae and surrounding soft tissue is very low [54], making identification of anatomical landmarks difficult [55]. Furthermore, although radiation dose for VF of the lumbar spine compares favourably with exposures for a single plain radiograph of the same region [55,56]; the risks associated with radiation exposure [57] remain present.

As mentioned previously, manual mark-up of VF images remains necessary [47] but differences in mark-up practices exist [56]. Currently, there is no consensus as to which is the most effective [56]. What is more, it is a laborious and time consuming process [54,55] that remains a source of error [48], and the choice of anatomical landmark identification can greatly influence the results [56]. Moreover, the optical distortion and out of plane motions [56] are likely to pose significant challenge to the clarity of VF images and ultimately, its usefulness in quantifying segmental ROM.

4.1.2 US

US imaging is a safe, inexpensive modality [58] which is portable, offering easy collection of static and dynamic images [59]. Though there are no known deleterious effects of US it remains the domain of competent sonographers [60].

Whilst it isn't commonplace to US image the spine, there is evidence that nearly all structures within the spine are visible with US [61]. However, despite adequate visualisation of structures being outlined by Ahmed et al. [61], the skill of completing US scanning largely remains operator dependent. For example, Margarido et al. [62] showed 20 unsupervised trials plus teaching sessions were not enough for participants to achieve competence in different aspects of US assessment of the lumbar spine. Therefore, if US imaging was to become more routine for assessing segmental ROM of the spine; specific training may be necessary. Furthermore, as US machines evolve, enhancements in image quality are further likely to facilitate easier imaging of the spine [61].

In comparison to other modalities, field of vision is small with US and directly limited to the area beneath the US transducer [63]. Also, distinct individual characteristics, such as BMI, are likely to affect the image quality; meaning this

modality may not be universally appropriate [63]. Despite this, real time analysis, video capture and enhancements to the technology and its image processing are likely solutions.

In summary, US is an inexpensive, safe and accessible modality that is already used extensively in clinical practice for other purposes. Therefore, it affords great potential for regular monitoring of lumbar spinal ROM. Nevertheless, it requires a skilled operator to image the lumbar spine and resolution of images may vary between patients based on extraneous patient variables or sonographer expertise.

4.1.3 MRI

MRI uses non-ionising radiation [64], is non-invasive [65] and is considered a safe technology [66]. Furthermore, it offers real advantages in terms of image quality, resolution and consistency [61,64]. MRI has the ability to visualise the entire spine, spinal cord and surrounding structures in its entire length [65]; providing further opportunities such as, the identification of structural changes. Moreover, MRI can produce sectional images of equivalent resolution, in any projection, without moving a patient [67]. This ability to obtain images in multiple planes adds to its versatility [67].

Analysis of spinal ROM requires the use of open MRI which eliminates a patients feeling of claustrophobia, along with the associated implications of this effect, commonly seen with traditional closed MRI scanners [68,69]. However, it does have some disadvantages. This is represented mostly by the use of a low field magnet; resulting in low signal to noise ratio and leading to reduced image quality compared with the more common high field magnet [69]. Equally, patients with pacemakers and certain ferromagnetic appliances cannot be imaged with MRI [67], and patient throughput is slow compared with other imaging modalities [67,69].

A further significant drawback to MRI is that the equipment is not only expensive to purchase, but also to maintain and operate [67]. Additionally, greater technological expertise is required for utilisation of MRI rather than most other imaging modalities [67]; highlighting important limitations.

Altogether, it is evident that MRI has good spinal visualising capabilities; coupled with consistency in image acquisition and interpretation. What is more, this modality does not pose a risk to most patients and offers the clinical advantage of looking at intervertebral disc deformation and soft tissue providing additional insights for patients with known pathologies. However, the substantial cost associated with using this technology indicates its lack of suitability for regular monitoring of lumbar segmental ROM.

4.1.4 Radiography

Radiography remains a cost effective spinal imaging method [70] and the equipment is widely available [71]. Compared to other imaging modalities, like MRI, usually performed in the recumbent position; radiographs can also be taken in different anatomical positions [70]. Nonetheless, there are no established guidelines for imaging the thoraco-lumbar spine with radiographs [70,72] and it is required to be performed in a specialised room [71]. There are also errors associated with distortion, magnification and positioning of individuals [71,73]. Furthermore, lots of heterogeneity exits in the methodology of radiographic segmental ROM measurements [72].

The most significant disadvantage to radiography though is its use of ionising radiation [73,74]. This is a known mutagen that can increase the risk of diseases such as cancer [75]. In addition, a higher beam energy is required due to the lumbar spines large x-ray attenuation and imaging of this area involves exposure to radiosensitive

reproductive organs [76]. These risks are an important consideration for repeated radiography examinations.

To summarise, however cost effective radiography remains, the errors linked to image capture and variability in image analysis, coupled with the risks associated with ionising radiation exposure; makes this imaging modality unsuitable for frequent assessment of lumbar spinal ROM.

5. Summary

This review has explored four potential modalities for segmental ROM assessment. All methods offer high reliability but the detail of the experimental design is critical to understand the magnitude of error associated with each. Such information will enable researchers and clinicians to make informed decisions regarding the correct modality for their particular situations. Also, the data here will provide a platform of the current state of the knowledge, from which developments or enhancements can be made to better determine methods of segmental ROM information. Once established, such methods may enable clinicians and researchers the opportunities to explore the fundamental principles underpinning LBP assessment and treatment practice.

6. Limitations

Eligible papers in this review shared a number of the same authors leading to a potential risk of bias. Likewise, as studies published in a non-English language and grey literature were excluded from this review, publication bias may be evident. Additionally, due to the heterogeneity of the literature retrieved, some synthesis was based on a relatively small number of papers, sometimes even single articles. Therefore, the generalisability of the findings may be limited.

7. Conclusion

This review has provided a contemporary systematic analysis of the literature related to the reliability of VF, US, MRI and Radiography modalities currently used for non-invasive measurements of segmental ROM in the lumbar spine in-vivo. Excellent reliability is seen in all modalities. However, VF is limited by radiation exposure, as is radiography, and there is a high cost associated with MRI. Additionally, both modalities are not routinely available. US offers potential for routine clinical use, with its low cost and widespread availability, which has the opportunity to provide a truly non-invasive and risk free method of measuring segmental ROM in individuals with LBP. Despite this, further research is necessary to determine whether US imaging yields truly consistent measurements of segmental ROM in the lumbar spine and whether this is also evident in within- and between-day repeated measures. If a method of segmental ROM assessment can be developed for routine clinical practice it could be a useful tool to evaluate abnormal segmental motion due to pain, spinal pathology or surgical intervention; signifying its potential value in the assessment, diagnosis and management of a variety of spinal related conditions.

8. Acknowledgements

No acknowledgements this manuscript.

9. Declaration of interest

None to report.

10. References

[1] Buchbinder R, Blyth FM, March LM, et al. Placing the global burden of low back pain in context. Best Pract Res Clin Rheumatol. 2013; 27(5): 575-589.

[2] Global Burden of Disease and Injury Incidence and Prevalence Collaborators. Global, regional and national incidence, prevalence and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017; 390(10100): 1211-1259.

[3] Maniadakis N, Gray A. The economic burden of back pain in the UK. Pain. 2000; 84(1): 95-103.

[4] Ramdas J, Jella V. Prevalence and risk factors of low back pain. Int J Adv Med. 2018; 5(5): 1120-1123.

[5] Maetzel A, Li L. The economic burden of low back pain: a review of studies published between 1996 and 2001. Best Pract Res Clin Rheumatol. 2002; 16(1): 23-30.

[6] Murray CJL, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: A systematic analysis for the global burden of disease study 2010. Lancet. 2012; 380(9859): 2197-2223.

[7] Reeves NP, Cholewicki J, Milner T, et al. Trunk antagonist co-activation is associated with impaired neuromuscular performance. Exp Brain Res. 2008; 188(3): 457-463.

[8] Hodges P, van den Hoorn W, Dawson A, et al. Changes in the mechanical properties of the trunk in low back pain may be associated with recurrence. J Biomech. 2009; 42(1): 61-66.

[9] Karayannis NV, Smeets RJ, van den Hoorn W, et al. Fear of movement is related to trunk stiffness in low back pain. PLoS One. 2013; 8(6): e67779.

[10] Kulig K, Powers CM, Landel RF, et al. Segmental lumbar mobility in individuals with low back pain: in vivo assessment during manual and self-imposed motion using dynamic MRI. BMC Musculoskelet Disord. 2007; 8(8): [10 p.]. doi:10.1186/1471-2474-8-8

[11] Golightly YM, Goode AP, Cleveland RJ, et al. FRI0598 relationship of joint hypermobility with low back pain and lumbar spine osteoarthritis: a cohort study. Ann Rheum Dis. 2016; 75: 659.

[12] Haxby Abbott J, Fritz JM, McCane B, et al. Lumbar segmental mobility disorders: comparison of two methods of defining abnormal displacement kinematics in a cohort of patients with non-specific mechanical low back pain. BMC Musculoskelet Disord. 2006; 7: 45.

[13] Trudelle-Jackson E, Fleisher LA, Borman N, et al. Lumbar spine flexion and extension extremes of motion in women of different age and racial groups: the WIN study. Spine. 2010; 35(16): 1539-1544.

[14] Fritz JM, Erhard RE, Hagen BF. Segmental instability of the lumbar spine. Phys Ther. 1998; 78(8): 889-896.

[15] Gercek E, Hartmann F, Kuhn S, et al. Dynamic angular three-dimensional measurement of multisegmental thoracolumbar motion in vivo. Spine. 2008; 33(21): 2326-2333.

[16] Rozumalski A, Schwartz MH, Wervey R, et al. The in vivo three-dimensional motion of the human lumbar spine during gait. Gait Posture. 2008; 28(3): 378-384.

[17] Park S-A, Fayyazi AH, Yonemura KS, et al. An in vivo kinematic comparison of dynamic lumbar stabilization to lumbar discectomy and posterior lumbar fusion using radiostereometric analysis. Int J Spine Surg. 2012; 1(6): 87-92.

[18] Dombrowski ME, Rynearson B, LeVasseur C, et al. ISSLS prize bioengineering science 2018: dynamic imaging of degenerative spondylolisthesis reveals mid-rang dynamic lumbar instability not evident on static clinical radiographs. Eur Spine J. 2018; 27(4): 752-762.

[19] Okawa A, Shinomiya K, Komori H, et al. Dynamic motion study of the whole lumbar spine by videofluoroscopy. Spine. 1998; 23(16): 1743-1749.

[20] Takayanagi K, Takahashi K, Yamagata M, et al. Using cineradiography for continuous dynamic-motion analysis of the lumbar spine. Spine. 2001; 26(17): 1858-1865.

[21] McGregor AH, Anderton L, Gedroyc WMW, et al. The use of interventional open MRI to assess the kinematics of the lumbar spine in patients with spondylolisthesis. Spine. 2002; 27(14): 1582-1586.

[22] Huang K-Y, Line R-M, Lee Y-L, et al. Factors affecting disability and physical function in degenerative lumbar spondylolisthesis of L4-5: evaluation with axially loaded MRI. Eur Spine J. 2009; 18(12): 1851:1857.

[23] Chleboun GS, Amway MJ, Hill JH, et al. Measurement of segmental lumbar spine flexion and extension using ultrasound imaging. J Orthop Sports Phys Ther. 2012; 42(10): 880-885.

[24] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021; 372: n71. doi:10.1136/bmj.n71

[25] NHLBI. Health Topics, Study Quality Assessment Tools: Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies [internet]. Bethesda (MD): National Heart, Lung and Blood Institute; [cited 2019 Dec 16]. Available from: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

[26] Haas M, Nyiendo J, Peterson C, et al. Interrater reliability of roentgenological evaluation of the lumbar spine in lateral bending. J Manipulative Physiol Ther. 1990; 13(4).

[27] Maigne J-Y, Lapeyre E, Morvan G, et al. Pain immediately upon sitting down and relieved by standing up is often associated with radiological lumbar instability or marked anterior loss of disc space. Spine. 2003; 28(12): 1327-1334.

[28] Breen AC, Muggleton JM, Mellor FE. An objective spinal motion imaging assessment (OSMIA): realiability, accuracy and exposure data. BMC Musculoskelet Disord. 2006; 7(1): [10 p.]. doi:10.1186/1471-2474-7-1

[29] Cakir B, Richter M, Käfer W, et al. Evaluation of lumbar spine motion with dynamic x-ray. Spine. 2006; 31(11): 1258-1264.

[30] Cakir B, Richter M, Puhl W, et al. Reliability of motion measurements after total disc replacement: the spike and the fin method. Eur Spine J. 2006; 15(2): 165-173.

[31] Landel R, Kulig K, Fredericson M, et al. Intertester reliability and validity of motion assessments during lumbar spine accessory motion testing. Phys Ther. 2008; 88(1): 43-49.

[32] Pearson AM, Spratt KF, Genuario J, et al. Precision of lumbar intervertebral measurements. Spine. 2011; 36(7): 572-580.

[33] Sui F, Zhang D, Lam SCB, et al. Auto-tracking system for human lumbar motion analysis. J Xray Sci Technol. 2011; 19(2): 205-218.

[34] Mellor FE, Thomas PW, Thompson P, et al. Proportional lumbar spine inter-vertebral motion patterns: a comparison of patients with chronic, non-specific low back pain and healthy controls. Eur Spine J. 2014; 23(10): 2059-2067.

[35] Yeager MS, Cook DJ, Cheng BC. Reliability of computer-assisted lumbar intervertebral measurements using a novel vertebral motion analysis system. Spine J. 2014; 14(2): 274-281.

[36] Cuesta-Vargas AI. Development of a new ultrasound-based system for tracking motion of the human lumbar spine: reliability, stability and repeatability during forward bending movement trials. Ultrasound Med Biol. 2015; 41(7): 2049-2056.

[37] Tozawa R, Katoh M, Aramaki H, et al. Reliability and validity of an ultrasound-based imaging method for measuring interspinous process distance in the lumbar spine using two different index points. J Phys Ther Sci. 2015; 27(7): 2333-2336.

[38] du Rose A, Breen A. Relationships between lumbar inter-vertebral motion and lordosis in healthy adult males: a cross sectional cohort study. BMC Musculoskelet Disord. 2016 [cited 2019 Oct 10]; [9 p.]. https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-016-0975-1

[39] Tozawa R, Katoh M, Aramaki, H, et al. Absolute and relative reliability of lumbar interspinous process ultrasound imaging measurements. J Phys Ther Sci. 2016; 28(8): 2210-2213.

[40] Breen A, Hemming R, Mellor F, et al. Intrasubject repeatability of in vivo intervertebral motion parameters using quantitative fluoroscopy. Eur Spine J. 2018; 28(2): 450-460.

[41] Mahato NK, Montuelle S, Clark BC. Assessment of in vivo lumbar inter-vertebral motion: reliability of a novel dynamic weight-bearing magnetic resonance imaging technique using a side bending task. Asian Spine J. 2019; 13(3): 377-385.

[42] Modica MJ, Kanal KM, Gunn ML. The obese emergency patient: imaging challenges and solutions. Radiographics. 2011; 31(3): 811-823.

[43] Cushman D, Flis A, Jensen B, et al. The effect of body mass index on fluoroscopic time and radiation dose during sacrioiliac joint injections. PM&R. 2015; 8(8): 767-772.

[44] Metaxas V, Messaris GA, Lekatou AN, et al. Patient dose in digital radiography utilising BMI classification. Radiat Prot Dosimetry. 2019; 184(2): 155-167.

[45] Durbridge G. Magnetic resonance imaging: fundamental safety issues. J Orthop Sports Phys Ther. 2011; 41(11): 820-828.

[46] Breen AC, Teyhen DS, Mellor FE, et al. Measurement of intervertebral motion using quantitative fluoroscopy: report of an international forum and proposal for use in the assessment of degenerative disc disease in the lumbar spine. Adv Orthop. 2012 [cited 2020 Jan 13]; [10 p.]. https://www.hindawi.com/journals/aorth/2012/802350/#B22

[47] Davis RJ, Lee DC, Wade C, et al. Measurement performance of a computer assisted vertebral motion analysis system. Int J Spine Surg. 2015;9: [13 p.]. doi:10.14444/2036

[48] Plocharski M, Lindstroem R, Lindstroem CF, et al. Motion analysis of the cervical spine during extension and flexion: Reliability of the vertebral marking procedure. Med Eng Phys. 2018; 61: 81-86.

[49] Stergiou N, Decker LM. Human movement variability, nonlinear dynamics and pathology: is there a connection?. Hum Mov Sci. 2011; 30(5): 869-888.

[50] Krüger M, Straube A, Eggert T. The propagation of movement variability in time: a methodological approach for discrete movements with multiple degrees of freedom. Front Comput Neurosci. 2017; 11: [11 p.]. doi:10.3389/fncom.2017.00093

[51] Yukawa Y, Matsumoto T, Kollor H, et al. Local sagittal alignment of the lumbar spine and range of motion in 627 asymptomatic subjects: age-related changes and sex-based differences. Asian Spine J. 2019; 13(4): 663-671.

[52] Galbusera F, Niemeyer F, Tao Y, et al. ISSLS prize in bioengineering science 2021: in vivo sagittal motion of the lumbar spine in low back pain patients-a radiological big data study. Eur Spine J. 2021 [cited 2021 Jan 25]; [9 p.]. doi: 10.1007/s00586-021-06729-z

[53] Manninen H, Kiekara O, Soimakallio S, et al. Reduction in radiation dose and imaging costs in scoliosis radiography. Application of large-screen intensifier photofluorography. Spine. 1988; 13(4): 409-412.

[54] Lam DCB, McCane, B, Allen R. Automated tracking in digitized videofluoroscopy sequences for spine kinematic analysis. Image Vis Comput. 2009; 27(10): 1555-1571.

[55] Muggleton JM, Allen R. Automatic location of vertebrae in digitilized videofluoroscopic images of the lumbar spine. Med Eng Phys. 1997; 19(1): 77-89.

[56] Harvey S, Hukins D, Smith F, et al. Measurement of lumbar spine intervertebral motion in the sagittal plane using videofluoroscopy. J Back Muculoskelet Rehabil. 2016; 29(3): 445-457.

[57] Kim HM, Choi KH, Kim TW. Patients' radiation dose during videofluoroscopic swallowing studies according to underlying characteristics. Dysphagia. 2013; 28(2): 153-158.

[58] Heidari P, Farahbakhsh F, Rostami M, et al. The role of ultrasound in diagnosis of the causes of low back pain: a review of the literature. Asian J Sports Med. 2015; 6(1): e23803.

[59] Marshburn TH, Hadfield CA, Sargsyan AE, et al. New heights in ultrasound: first report of spinal ultrasound from the international space station. J Emerg Med. 2014; 46(1): 61-70.

[60] The World Federation for Ultrasound in Medicine and Biology (WFUMB). WFUMB policy and statements on safety of ultrasound. Ultrasound Med Biol. 2013; 39(5): 926-929.

[61] Ahmed AS, Ramakrishnan R, Ramachandran V, et al. Ultrasound diagnosis and therapeutic intervention in the spine. J Spine Surg. 2018; 4(2): 423-432.

[62] Margarido CB, Arzola C, Balki, M, et al. Anaesthesiologists' learning curves for ultrasound assessment of the lumbar spine. Can J Anaesth. 2010; 57(2): 120-126.

[63] Hides JA, Richardson CA, Jull GA. Use of real-time ultrasound imaging for feedback in rehabilitation. Man Ther. 1998; 3(3): 125-131.

[64] Wassenaar M, van Rijn RM, van Tulder MW, et al. Magentic resonance imaging for diagnosing lumbar spinal pathology in adult patient with low back pain or sciatica: a diagnostic systematic review. Eur Spine J. 2012; 21(2): 220-227.

[65] Sett P, Crockard HA. The value of magnetic resonance imaging (MRI) in the follow-up management of spinal injury. Paraplegia. 1991; 29(6): 396-410.

[66] Hartwig V, Giovannetti G, Vanello N, et al. Biological effects and safety in magnetic resonance imaging: A review. Int J Environ Res Public Health. 2009; 6(6): 1778-1798.

[67] National Heart Lung and Blood Institute. Laboratory of Cardiac Energetics: What are the technological advantages and limitations (disadvantages) of MRI [internet]. Bethesda (MD): National Heart, Lung and Blood Institute; [cited 2020 Mar 20]. Available from: https://dir.nhlbi.nih.gov/labs/lce/cmri/mri-advantages-limitation.asp

[68] Tarantino U, Fanucci E, lundusi R, et al. Lumbar spine MRI in upright position for diagnosing acute and chronic low back pain: statistical analysis of morphological changes. J Orthop Taumatol. 2013; 14(1): 15-22.

[69] Michelini G, Corridore A, Torlone S, et al. Dynamic MRI in the evaluation of the spine: state of the art. Acta Biomed. 2018; 89: 89-101.

[70] Oakley PA, Harrison DE. Radiophobia: 7 reasons why radiography used in spine posture rehabilitation should not be feared or avoided. Dose-Response. 2018; 16(2): 1559325818781445.

[71] Janssen M, Nabih A, Moussa W, et al. Evaluation of diagnosis techniques used for spinal injury related back pain. Pain Res Treat. 2011 [cited 2021 Jan 27]; [10 p.]. https://www.hindawi.com/journals/prt/2011/478798/

[72] Leone A, Guglielmi G, Cassar-Pullicino VN, et al. Lumbar intervertebral instability: A review. Radiology. 2007; 245(1): 62-77.

[73] Mellor FE, Thomas P, Breen A. Moving back: radiation dose received from lumbar spine quantitative fluoroscopy compared to lumbar spine radiographs with suggestions for dose reduction. Radiography. 2014; 20(3): 251-257.

[74] Davey E, England A. AP versus PA positioning in lumbar spine computed radiography: image quality and individual organ doses. Radiography. 2014; 21(2): 188-196.

[75] Logan GS, Pike A, Copsey B, et al. What do we really know about the appropriateness of radiation emitting imaging for low back pain in primary emergency care? A systematic review and meta-analysis of medical record reviews. PLoS ONE. 2019; 14(12): e0225414.

[76] Lai ZH, Sá dos Reis C, Sun Z. Effective dose and image optimisation of lateral lumbar spine radiography: a phantom study. Eur Radiol Exp. 2020; 4(13) : [9 p.]. doi:10.1186/s41747-019-0132-3

Table 1. Search strategy

Search Strategy	Number of results (with duplicates removed)
Fluoroscopy AND spine AND motion AND reliability OR validity OR consistency OR repeatability NOT cervical OR thoracic	41
Ultrasound OR ultrasonography OR US OR USS OR ultrasound imaging AND lumbar AND reliability OR validity OR consistency OR repeatability NOT scoliosis OR musc* OR cervical OR thoracic	157
Magnetic resonance imaging OR MRI OR MRI scan AND lumbar spine AND motion OR kinematics AND reliability OR validity OR consistency OR repeatability	32
X ray OR radiology OR radiograph* AND reliability OR validity OR consistency OR repeatability AND lumbar AND motion OR kinematics NOT videofluoroscopy OR fluoroscopy OR scoliosis	200
Total	430

(* (asterisk), truncation).

Table 2. Detailed inclusion and exclusion criteria

Inclusion Criteria	Exclusion criteria
Human participants	Animal studies/ studying in-vitro
Measuring segmental or intersegmental ROM defined as angular rotation of one vertebral body on another (or a representation of this)	Articles solely investigating linear translation of a vertebrae
Measurements at the lumbar spine	Measuring ROM in only cervical and/ or thoracic spine
Measuring ROM with VF, US, MRI or radiography	Not using modality for imaging
Investigating reliability or validity	Studies published after January 2021
	Non-objective psychometric outcome
	Studies published not in the English language

(ROM, range of motion; VF, video fluoroscopy; US, ultrasound; MRI, magnetic resonance imaging).

Table 3. Reason for article rejection after accessing full-text citation

Reason for full-text citation rejection	No. of citations in this category
Not examining modality's psychometric properties	7
Examination in-vitro	4
Examination of an animal spine	1
Not VF/ US imaging/ MRI/ radiography	2
No reliability statistics included	2
Not assessing kinematics, motion or ROM	1
Measuring only linear translation	4
A report	1
Comparing a method against modality	1
Total no. of citations rejected	23

(No., number of; ROM, range of motion; VF, video fluoroscopy; US, ultrasound; MRI, magnetic resonance imaging).

Table 4. Quality assessment

	Study																
Criterion	Haas et al. 1990 [26]	Maigne et al. 2003 [27]	Breen et al. 2006 [28]	Cakir et al. 2006 [29]	Cakir et al. 2006 [30]	Landel et al. 2008 [31]	Pearson et al. 2011 [32]	Sui et al. 2011 [33]	Chleboun et al. 2012 [23]	Mellor et al. 2014 [34]	Yeager et al. 2014 [35]	Cuesta-Vargas 2015 [36]	Tozawa et al. 2015 [37]	Du Rose and Breen 2016 [38]	Tozawa et al. 2016 [39]	Breen et al. 2018 [40]	Mahato et al. 2019 [41]
1	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
2	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y
3	CD	Y	CD	CD	CD	CD	N	CD	CD	Y	Y	CD	CD	CD	CD	CD	CD
4	Y	Y	N	Y	Y	Y	Y	N	CD	Y	Y	Y	CD	Y	CD	Y	CD
5	N	N	N	N	N	N	N	N	N	Y	N	Y	N	Y	N	N	N
6	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
9	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
10	N	N	N	Y	Y	N	N	N	N	N	N	N	N	Y	N	Y	Y
11	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
12	NA	NA	NA	NA	NA	NA	NA	NA	NA	N	N	NA	NA	NA	NA	NA	NA
13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NR	NA	NR	Y
14	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	Y	N
Overall	Fair	Good	Fair	Good	Good	Good	Good	Fair	Fair	Good	Good	Good	Fair	Good	Fair	Good	Good

(Y, yes; N, no; CD, cannot determine; NA, not applicable; NR, not reported; overall, overall quality rating).

Table 5. Data extraction table

Author	Modality	Participant characteristics	Psychometric outcome(s) assessed	Vertebral level(s)	Measurement Methods
Haas et al. 1990 [26]	Radiography	58 participants Mean age 28 Students from Western States Chiropractic College Asymptomatic	Kappa SEM	L1-L5	Investigated tilt into side bending and rotation in standing and lateral bending positions.
Maigne et al. 2003 [27]	Radiography	74 participants with Chronic LBP 42 had pain immediately on sitting down and relieved on standing up from sitting (mean age 54.9, 6 males, 36 females) 32 age and gender matched participants who did not have symptom described above (mean age 57.5, 4 males, 28 females)	Mean difference LoA	L1-L5	Investigated angle change between adjacent vertebral endplates of lateral flexion-extension radiographs in standing and sitting positions.
Breen et al. 2006 [28]	VF	30 male participants Aged 18-40 Asymptomatic 4 subjects assessed	RMSE	L3-L5	Investigated non weight bearing side flexion, flexion and extension. One movement trial for each individual. Manual identification of first frame then automated analysis using vertebral corners as reference points.

Cakir et al. 2006 [29]	Radiography	24 participants. 10 males, 14 females All with monosegmental degenerative disc disease Mean age 40.2	PCC	L4-S1	Investigated flexion-extension radiographs in standing. Three examiners, two took measurements between-day.
Cakir et al. 2006 [30]	Radiography	24 participants. 10 males, 14 females following a monosegmental total disc replacement at L4-5 or L5-S1 Mean age 40.2	PCC	L4-S1	Investigated flexion-extension radiographs in standing. Follow up study. Three examiners, two took measurements between-day.
Landel et al. 2008 [31]	MRI	29 participants 13 Males, 16 Females Aged 18-45 Diagnosis of non-specific LBP Recent onset of centralised LBP	ICC SEM	L1-L5	Investigated P-A force in non-weight bearing. Segmental mobility quantified by measuring the change in the intervertebral angle between the resting position and the end range of the P-A force application.
Pearson et al. 2011 [32]	Radiography	30 participants with spondylolisthesis at L4-5 37% males, 63% females Randomly selected from the spine patient outcomes research trial Mean age 66	ICC SEM	L1-S1	Investigated intervertebral rotation of flexion-extension radiographs. Measurements made with a digitised manual technique by three raters and by a quantitative motion analysis software by three different raters.
Sui et al. 2011 [33]	VF	12 participants 10 healthy, 2 lumbar spondylolisthesis 8 Males, 4 Females	ICC	L1-S1	Investigated seated flexion and extension. Automated tracking after manual marking of four vertebral corners.

		Aged 19-38			
Chleboun et al. 2012 [23]	MRI US	6 participants 2 Males, 4 Females Aged 22-35 Asymptomatic	ICC CoV	L1-L5	Investigated supine flexion and extension postures. MRI – distance between inferior edge of caudal and cranial spinous processes measured. US – distance between the peak of the curvature of caudal and cranial spinous processes measured. Manual/visual method used digitally.
Mellor et al. 2014 [34]	VF	80 participants 44 Males, 36 Females Aged 21-50 40 with Chronic non-specific LBP 40 asymptomatic volunteers Convenience sample	ICC SEM	L2-L5	Investigated lying flexion and extension. Sequences processed using automatic tracking algorithms after manual template application to first image. Landmark used was vertebral body corners.
Yeager et al. 2014 [35]	VF	61 participants 52% Male, 48% Female Aged 31-62 34 Asymptomatic 14 preoperative (with confirmed pathology) 13 post operative (with a previous lumbar procedure)	ICC SEM	L1-S1	Investigated flexion and extension in upright and lying positions. Intervertebral rotation and intervertebral translation measured using automated vertebral motion analysis tracking algorithms as well as a manual technique.

Cuesta-Vargas 2015 [36]	US	15 male participants Convenience sample Asymptomatic	ICC SEM MDC	L4-L5	Investigated flexion in upright position. 10 reps forward bending and neutral. Manual/visual identification of spinous process then semi-automated orientation estimation.
Tozawa et al. 2015 [37]	US	10 participants Healthy Males Aged 20-23	ICC MDC	L2-L3	Investigated prone, prone on elbows and kneeling with flexed spine postures. Measured distance between caudal end of L2 spinous process to cranial end of L3 spinous process and from top of L2 spinous process to top of L3 spinous process. Manual/ visual identification of landmarks.
du Rose and Breen 2016 [38]	VF	18 male participants Mean age 27.6 No history of LBP	ICC SEM	L2-S1	Investigated flexion and extension in upright position. Sequences processed using automatic tracking algorithms after manual template application to first image.
Tozawa et al. 2016 [39]	US	10 male participants Aged 20-23 No history of orthopaedic disease or dysfunctions	ICC MDC	L1-L2	Investigated prone, prone on elbows and kneeling with Lx fully flexed postures. Manual/ visual identification of spinous process. Measured distance between caudal end of L1 and cranial end of L2 spinous processes.

Breen et al. 2018 [40]	VF	109 participants 66 Males, 43 Females Aged 21-80 Healthy volunteers Convenience sample	ICC MDC	L2-S1	Investigated flexion, extension, left side flexion and right side flexion in recumbent and standing positions. Single motion sequences. Manual first image registration then frame-to-frame automatic tracking.
Mahato et al. 2019 [41]	MRI	10 participants Aged 18-60 Volunteers Asymptomatic	ICC CoV	L2-L4	Investigated side flexion in weight-bearing upright position. Measured changes in intervertebral axes positions using cranial to caudal vertebrae measurement and displacements in individual vertebrae within a calibrated imaging space.

(SEM, standard error of measurement; L, lumbar; LoA, limits of agreement; VF, videofluoroscopy; RMSE, root mean square error; PCC, pearsons correlation coefficient; MRI, magnetic resonance imaging; LBP, lower back pain; ICC, intra-class correlation co-efficient; P-A, posterior-anterior; %, percent; US, ultrasound; CoV, coefficient of variation; MDC, minimal detectable change; Lx, Lumbar spine).

Table 6. Data extraction findings

Author		
	Video fluoroscopy	
Breen et al. 2006 [28]	Side bending	Intra-subject variation 2.75º RMSE (observer 1) Intra-subject variation 2.91º RMSE (observer 2) Raw data ranges not reported
Sui et al. 2011 [33]		Did not report individual segmental ROM values Only reported ICC for the plastic spine model
Mellor et al. 2014 [34]	Segmental ROM	
	Flexion	L2/3 = 4.23º, L3/4 = 5.89º, L4/5 = 7.10º (Patients) L2/3 = 4.05º, L3/4 = 5.49º, L4/5 = 6.46º (Controls)
	Extension	L2/3 = 5.04º, L3/4 = 4.15º, L4/5 = 4.78º (Patients) L2/3 = 4.64º, L3/4 = 4.11º, L4/5 = 5.31º (Controls)
	Intra-observer	
	Flexion	SEM% = 3% L2/3 SEM 0.13º, ICC 0.98, CI95% 0.86-0.99 L3/4 SEM 0.13º, ICC 0.98, CI95% 0.90-1.0 L4/5 SEM 0.10º, ICC 1.0, CI95% 0.99-1.0
	Extension	SEM% = 8% L2/3 SEM 0.35º, ICC 0.96, CI95% 0.85-0.99 L3/4 SEM 0.24º, ICC 0.92, CI95% 0.72-0.98 L4/5 SEM 0.19º, ICC 0.99, CI95% 0.97-1.0
	Side bending	SEM % = 3%. SEM 0.08-0.17º, ICC 0.99-1.0, CI95% 0.95-1.0
	Inter-observer	
	Flexion	ICC 0.74-0.99; CI95% 0.23-0.99 SEM 0.17-0.31º SEM% = 7% L2/3 SEM 0.31º, ICC 0.91, CI95% 0.69-0.98 L3/4 SEM 0.17º, ICC 0.98, CI95% 0.91-0.99 L4/5 SEM 0.31º, ICC 0.97, CI95% 0.88-0.99
	Extension	SEM% = 19%

		L2/3 SEM 0.77º, ICC 0.76, CI95% 0.27-0.94 L3/4 SEM 0.41º, ICC 0.74, CI95% 0.23-0.93 L4/5 SEM 0.27º, ICC 0.99, CI95% 0.96-1.00
	Side bending	SEM 0.18-0.55º, ICC 0.85-0.99, CI95% 0.51-1.00
Yeager et al. 2014 [35]	Segmental ROM Flexion Extension	 4.40º 2.00º
	Intra-observer Flexion/Extension	 SEM 0.10º, ICC 0.98, CI95% 0.98-0.99 SEM% = 2% (flexion) SEM% = 5% (extension)
	Inter-observer Flexion/Extension	 SEM 0.22º, ICC 0.96, CI95% 0.95-0.97 SEM% = 5% (flexion) SEM% = 11% (extension)
du Rose and Breen 2016 [38]	Between day Intra-rater reliability SEM	L2/3 SEM 0.45º, ICC CI95% 0.92-1.0 L3/4 SEM 0.23º, ICC CI95% 0.96-1.0 L4/5 SEM 0.39º, ICC CI95% 0.97-1.0 L5/S1 SEM 0.54º, ICC CI95% 0.82-0.99
Breen et al. 2018 [40]	Segmental ROM Lying flexion Lying extension Standing flexion Standing extension	 5.14º 5.33º 9 º - 14º 2.01º
	Repeated measures Lying flexion / extension Standing flexion / extension Lying side bending	 ICC 0.80, CI95% 0.74-0.85 MDC_{95} 4.66º, MDC% 91% (flexion) MDC_{95} 5.19º, MDC% 97% (extension) ICC 0.82-0.91, CI95% 0.76-0.93 MDC_{95} 9.10º, MDC% 100% (flexion) MDC_{95} 5.53º, MDC% 176% (extension) ICC 0.95, CI95% 0.92-0.96 MDC_{95} 3.3-3.7º, MDC% 60-69%

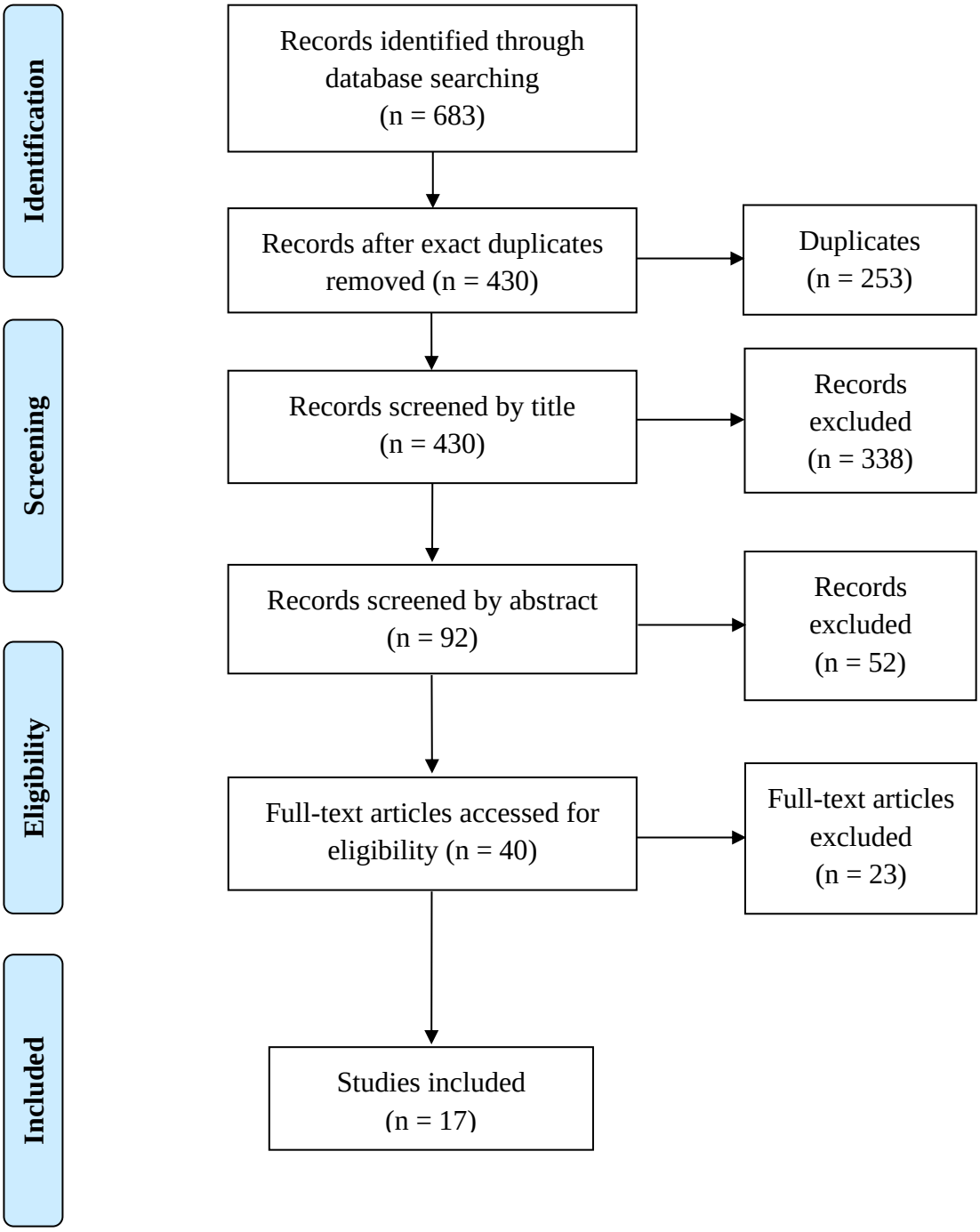
	Standing side bending	ICC 0.90-0.92, CI95% 0.0.87-0.94 MDC_{95} 4.5-4.7º, MDC% 97-98%
	Ultrasound	
Chleboun et al. 2012 [23]	Segmental ROM Flexion Extension	 Spinous process distance = 25.6mm to 32.3mm ROM taken from neutral = 3.0mm to 4.4mm Spinous process distance = 21.5mm to 26.9mm
	Reliability	ICC 0.94, CI95% 0.85-0.97 CoV = 1.8%
Cuestas-Vargas 2015 [36]	Segmental ROM Flexion	 15.5º +/- 2.04º
	Repeated measures	ICC CI95% = 0.995-0.999 (within day) CI95% = 0.996-0.999 (between day) MDC 1.5º (or 10%).
Tozawa et al. 2015 [37]	Segmental ROM	Lumbar interspinous process distance ranged from 29.2mm to 30.1mm
	Intra-rater reliability	ICC CI95% 0.79-1.0 Examiner A: ICC 0.97-0.98, CI95% 0.93-0.99 Examiner B: ICC 0.96-0.98, CI95% 0.92-0.99 Examiner C: ICC 0.97-0.98, CI95% 0.94-0.99 Examiner D: ICC 0.97-0.99, CI95% 0.94-1.0 Examiner E: ICC 0.90-0.99, CI95% 0.79-1.0 MDC_{95} value of 0.29mm
	Inter-rater reliability	ICC 0.914, CI95% 0.80-0.97 ICC 0.725, CI95% 0.55-0.87
Tozawa et al. 2016 [39]	Intra-rater reliability	ICC 0.990-0.998, CI95% 0.963-0.999. Measurer A: ICC 0.997, CI95% 0.993-0.999. Measurer B: ICC 0.992, CI95% 0.981-0.998. Measurer C: ICC 0,998, CI95% 0.996-0.999. Measurer D: ICC 0.985, CI95% 0.963-0.996. Measurer E: ICC 0.991, CI95% 0.978-0.997. Measurer F: ICC 0.995, CI95% 0.987-0.998. Measurer G: ICC 0.995, CI95% 0.987-0.999. Measurer H: ICC 0.992, CI95% 0.980-0.998. Measurer I: ICC 0.990, CI95% 0.977-0.997. MDC_{95} values of 0.62-1.8mm. Measurer D: MDC_{95} = 1.8mm. Measurer F: MDC_{95} = 1.1mm. Measurer H: MDC_{95} = 0.62mm. Measurer I: MDC_{95} = 1.5mm
	Inter-rater reliability	ICC 0.969, CI95% 0.90-1.0

	Magnetic resonance imaging	
Landel et al. 2008 [31]	Reliability	ICC 0.95-0.99 SEM ranged from 0.40º to 0.66º
Chleboun et al. 2012 [23]	Segmental ROM Flexion Extension	 Spinous process distance = 24.6mm to 35.6mm ROM estimates from neutral = 1.8mm to 4.9mm Spinous process distance = 19.9mm to 29.4mm ROM estimates from neutral = 0.9mm to 4.3mm
	Reliability	ICC 0.98, CI95% = 0.95-0.99. CoV = 1.6%.
Mahato et al. 2019 [41]	Segmental ROM Side bending (right)	 8.5º to 17.3º
	Between day Repeated measures	 ICC 0.93-0.94 CoV 14-15%
	Radiography	
Haas et al. 1990 [26]	Tilt Neutral Left lateral bending Right lateral bending	Kappa L1 = 0.47, L2 = 0.56, L3 = 0.46, L4 = 0.22, L5 = 0.17 L1 = 0.50, L2 = -0.03, L3 = 1.00, L4 = 0.25, L5 = 0.19 L1 = 0.24, L2 = 0.25, L3 = 0.00, L4 = 0.27, L5 = 0.16
	Rotation Neutral Left lateral bending Right lateral bending	Kappa L1 = 0.64, L2 = 0.68, L3 = 0.55, L4 = 0.60, L5 = 0.57 L1 = 0.63, L2 = 0.61, L3 = 0.57, L4 = 0.55, L5 = 0.42 L1 = 0.49, L2 = 0.42, L3 = 0.59, L4 = 0.68, L5 = 0.38
	SEM (one rater pair) Neutral radiographs Net tilt Net rotation	 1.0º - 1.9º (tilt); 2.1º - 2.6º (rotation) 1.3º - 3.2º (left lateral bending) 1.2º - 2.8º (right lateral bending) 2.0º - 3.7º (left lateral bending) 2.5º - 3.6º (right lateral bending)
Maigne et al. 2003 [27]	Angular motion Extension to flexion	 13.9º ± 4.5º (patient group) 7.5º ± 4.3º (control group)
	Extension to sitting	10.0º ± 4.5º (patient group) 6.2º ± 4.0º (control group)
	Intra-rater (mean difference between end plate angle measurements)	0.31º, 95%CI LoA -3.0º to 2.4º (extension) 0.04º, 95%CI LoA -3.0º to 3.0º (flexion) 0.03º, 95%CI LoA -3.0º to 3.0º (sitting)

	Inter-rater (mean difference between end plate angle measurements)	-0.38º, 95%CI LoA -3.1º to 3.9º (extension) -0.44º, 95%CI LoA -2.7º to 3.6º (flexion) -1.05º, 95%CI LoA -2.7º to 4.8º (sitting)
Cakir et al. 2006 [29]	Intra-rater reliability	PCC (95%CI) = 0.902 (±4.2º), 0.782 (±6.8º), 0.916 (±4.0º), 0.881 (±4.7º). Mean difference = -0.17º, 0.04º, -0.17º, -0.17º.
	Inter-rater reliability	Range of PCC (95%CI) = 0.843, 0.809, 0.777, 0.738 (-7.4º/ +5.8º); 0.929, 0.913 (-4.5º/ +4.3º); 0.890, 0.861, 0.890, 0.891 (-4.9º/ +4.5º); 0.885, 0.888 (-5.0º/ +4.2º). Mean difference between 2 measurement sets = -0.82º, -0.07º, -0.17º, -0.38º
Cakir et al. 2006 [30]	Intra-rater reliability	PCC (95%CI) = 0.962 (±2.1º), 0.903 (±3.3º), 0.955 (±2.0º), 0.916 (±3.0º) Mean difference = 0.04º, 0.08º, -0.08º, -0.04º
	Inter-rater reliability	Range of PCC (95%CI) = 0.928, 0.903, 0.911, 0.917 (-3.0º/ +3.0º); 0.918, 0.905 (-2.9º/ +3.1º); 0.899, 0.930, 0.950, 0.950 (-2.4º/ +3.0º); 0.926, 0.886 (-2.8º/ +2.8º) Mean difference between 2 measurement sets = 0.04º, -0.06º, -0.31º, -0.00º
Pearson et al. 2011 [32]	Segmental ROM	Average change in intervertebral rotation = 5.1º (DMT), 5.7º (QMA).
	Intra-rater reliability	ICC = 0.870 SEM = 2.5º (DMT) ICC =0.997 SEM = 0.5º (QMA)
	Inter-rater reliability	ICC = 0.693 (DMT) ICC = 0.976 (QMA)

(L, lumbar; =, equals; SEM, standard error of measurement; º, degrees; 95%CI, ninety five percent confidence interval; LoA, limits of agreement; RMSE, root mean square error; PCC, pearson correlation coefficient; ±, plus or minus; ICC, intra-class correlation co-efficient; DMT, digitised manual technique; QMA, quantitative motion analysis; US, ultrasound; mm, millimetres; ROM, range of motion; CI95%, ninety five percent confidence interval; CoV, coefficient of variation; MRI, magnetic resonance imaging; L2/3, lumbar spine intervertebral level 2/3; L3/4, lumbar spine intervertebral level 3/4; L4/5, lumbar spine intervertebral level 4/5; SEM%, standard error of measurement percent; %, percent; L5/S1, spine intervertebral level lumbar 5/ sacral 1; MDC_{95}, minimal detectable change at 95% confidence level; MDC%, percentage magnitude of minimal detectable change at 95% confidence level).

Figure 1. PRISMA flow diagram of search



(n, number; =, equals).