Medical Engineering and Physics 126 (2024) 104146

Contents lists available at ScienceDirect

Medical Engineering and Physics

journal homepage: www.elsevier.com/locate/medengphy

Review

Validity and reliability of inertial measurement units used to measure motion of the lumbar spine: A systematic review of individuals with and without low back pain

Frederick A. McClintock [*], Andrew J. Callaway, Carol J. Clark, Jonathan M. Williams

Faculty of Health and Social Sciences, Bournemouth University, Fern Barrow, Poole BH12 5BB, United Kingdom

ARTICLE INFO

Keywords:
Low back
Pain
Sensors
Review
Biomechanics

ABSTRACT

Low back pain (LBP) is a leading cause of disability, resulting in aberrant movement. This movement is difficult to measure accurately in clinical practice and gold standard methods, such as optoelectronic systems involve the use of expensive laboratory equipment. Inertial measurement units (IMU) offer an alternative method of quantifying movement that is accessible in most environments. However, there is no consensus around the validity and reliability of IMUs for quantifying lumbar spine movements compared with gold standard measures. The aim of this systematic review was to establish concurrent validity and repeated measures reliability of using IMUs for the measurement of lumbar spine movements in individuals with and without LBP. A systematic search of electronic databases, incorporating PRISMA guidelines was completed, limited to the English language. 503 studies were identified where 15 studies met the inclusion criteria. Overall, 305 individuals were included, and 109 of these individuals had LBP. Weighted synthesis of the results demonstrated root mean squared differences of <2.4° compared to the gold standard and intraclass correlations >0.84 for lumbar spine movements. IMUs offer clinicians and researchers valid and reliable measurement of motion in the lumbar spine, comparable to laboratory methods, such as optoelectronic motion capture for individuals with and without LBP.

1. Introduction

Low back pain (LBP) is the leading cause of years lived with disability [1,2]. Prevalence estimates vary across reports, but range between 23 and 42 % of the global population [3,4]. Despite initial improvements to pain within 12 weeks of an acute episode of LBP, a review of 11 studies found up to 71 % of individuals report ongoing symptoms [5].

The measurement of lumbar motion remains a common part of impairment evaluation in the assessment of individuals with LBP. Clinically, LBP is assessed using separate, single, in-plane movements for flexion/extension, lateral bending, or rotation [6] where measurement methods commonly rely on observation [7], a tape measure [8], or an inclinometer [9]. Whilst quick and easy to implement, observation cannot quantify the movement, whilst a tape measurement and inclinometer only quantify the peak angle achieved at the end of the range of movement. The inability to quantify continuous lumbar angle throughout a single-plane movement, or through complex dynamic movements in multiple planes of motion, such as lunge and twist movements, or movements in sport such as a cricket fast bowling, limits the understanding of lumbar kinematics through time.

Laboratory systems often utilise optoelectronic motion capture to measure movement through time, and these are thought to be the gold standard non-invasive method of measuring kinematics [10,11]. However, these are often limited by cost, complexity, are time-consuming, and are usually constrained to a dedicated environment [12,13]. Through-movement analysis is required to allow movement quality to be assessed [14]. Therefore, researchers and clinicians are seeking alternative methods to provide valid and reliable objective motion analysis that overcomes these limitations and moves the understanding of LBP movement forward.

One potential solution is miniature body-worn sensors such as inertial measurement units (IMUs) [15]. These devices house tri-axial accelerometers, gyroscopes, and magnetometers and through the fusion of these elements, offer drift-free orientation data, from which joint angles can be derived [16,17]. The growth of the use of such devices to measure joint angles is evident from the literature [18–20], however, previous authors have suggested that the validity and reliability of such devices applied to human motion analysis may be dependent on the specific

---

[*] Corresponding author.
E-mail address: fmcclintock@bournemouth.ac.uk (F.A. McClintock).

https://doi.org/10.1016/j.medengphy.2024.104146
Received 6 October 2023; Received in revised form 22 December 2023; Accepted 9 March 2024

anatomical region being assessed, as well as the task being performed [10]. Therefore, it is essential to understand if IMUs can offer a viable solution for motion analysis of the lumbar spine.

Prior to recommending a new measurement method, establishing validity and reliability is critical. As individuals with LBP move differently to those without [21–24], and because clinical studies often compare individuals with matched controls, it is necessary to establish validity, and reliability in individuals with and without LBP. IMUs are known to perform differently to slow and fast movements [25,26], therefore establishing validity and reliability in both groups is important. This evaluation is important for understanding the devices' sensitivity, and discriminatory ability in accurately measuring movement behaviour associated with underlying LBP.

Previous reviews focussing on concurrent validity and reliability have demonstrated IMUs are valid and reliable for measuring stride variables during running and gait analysis [27,28]. Moreover, a previous review has explored the state of the art of wearables for spinal movement testing [29], however the focus of this review was not IMUs specifically nor did it focus on validity and reliability. Therefore, to date no systematic evaluation specifically of validity and reliability for using IMUs to measure lumbar movement exists. If validity and reliability can be established, new opportunities for real world monitoring of lumbar movement will be possible. Therefore, the aim of this review is to establish concurrent validity (compared to a gold standard) and the repeated measures reliability of using IMUs for measurement of the movement of the lumbar spine, in individuals with and without LBP.

2. Methods

2.1. Search strategy

A search strategy was developed in line with PRISMA guidelines [30]. A Boolean search on databases MEDLINE, CINAHL, SPORTDiscuss, Scopus, and Web of Science, was performed in April 2022. A date limiter of 2005 was set as due to the rapidly evolving IMU sensor industry, ensuring only the most up-to-date measurement methods were considered. This allows the results of this review to be applicable for modern and current IMU devices. The keywords used for this search are presented in Table 1. Once completed, and with duplicates removed, the articles were analysed against inclusion and exclusion criteria (Section 2.2), with any uncertainty resolved by consensus (Authors: 1,2,3,4). Reference lists of included articles were reviewed with additional appropriate articles identified and further screened against inclusion and exclusion criteria. A diagram of the search can be seen in Fig. 1.

2.2. Inclusion/exclusion criteria

To be included, articles needed to have employed two or more IMUs to measure lumbar movement. This was to ensure studies were investigating the relative angle between two sensors, not the absolute angle of one sensor (including the use of a smartphone). As the focus of the review is on IMUs, articles using only accelerometers or gyroscopes in isolation were excluded. The articles needed to report either the concurrent validity of IMUs compared to optoelectronic motion capture or repeated measures reliability of IMUs, or both. The participants included in the articles needed to be healthy or have non-specific low back pain.

Table 1
Search strategy terms.

Boolean function	Search terms
	lumbar OR spine OR spinal OR "low spine" OR "lumbar spine" OR "low back" OR "lower back" OR "low-back" OR "lower-back"
AND	pain OR injury OR "low-back-pain"
AND	"inertial measurement unit*" OR IMU OR "inertial sensor" OR node

Articles reporting use in individuals with LBP in association with a known pathological condition were excluded, to minimise the risk of condition-specific bias. To be included in the quantitative synthesis the articles needed to include data values in tables, not just graphical formats. The data needed to be measuring low back movement during common clinical assessment movements (flexion extension, lateral flexion, rotation), or dynamic complex movements (involving movements in all three planes) such as sporting movements e.g., cricket fast bowling, discus throwing.

2.3. Study appraisal and data extraction

Studies that passed the inclusion requirements were subject to an appraisal and risk of bias check. To appraise studies reporting on validity, the QUADAS-2 checklist was used [31] and for studies reporting reliability, the QAREL checklist was used [32]. Studies that reported both validity and reliability were appraised using both checklists. Each article was assessed by the same author (FM) with uncertainties resolved by consensus (FM, JW). The results of the quality appraisal are found in Tables 2 and 3. Validity studies, in Table 2, gained a point for every "low risk" answer and due to the relevance of the questions, needed to score at least four to be included in the review. Reliability studies, in Table 3, scored a point for every "yes" answer and due to the relevance of the questions needed to score at least five points to be included in the review. These thresholds provided a consensus that the papers had low risk of bias throughout. All the studies passed the appraisal and risk of bias checks. Data from all the included studies were entered into a spreadsheet by a single investigator (FM) and checked by a second investigator (JW). This spreadsheet tabulated author names and publication date, study design, participants, technology and methods, lumbar spine assessment procedure, validity and/or reliability measures, statistical analysis, results, and any methodological limitations and comments (see Table 4). Further measures of reliability such as standard error measurement (SEMs) produced by some of the studies can be found in Table 5.

2.4. Quantitative synthesis

For the quantitative synthesis, validity studies needed to report root mean squared error (RMSE) and reliability studies intraclass correlations (ICC). The quantitative synthesis was performed for the three planes of motion, flexion extension, lateral flexion, and rotation. Complex movements (which involved movements in all three planes of movement) were compared against each other to attain a further understanding of IMU performance. Weighted mean, computed from the proportion of each study's sample size relative to the total cumulative sample size, with a 95 % confidence interval (95 %CI) were calculated from the IMU measurements for each of the three main movements for validity and reliability. The standard deviation of the RMSE from each study was combined using Eq. (1) [33]. To calculate the weighted standard deviation for reliability studies reporting ICC values, Eq. (2) was used.

$$\sqrt{\frac{(N_1-1)SD_1^2+(N_2-1)SD_2^2+\frac{N_1N_2}{N_1+N_2}\left(M_1^2+M_2^2-2M_1M_2\right)}{N_1+N_2-1}} \tag{1}$$

Where N_1 is the sample size for study 1 and N_2 is the sample size for study 2. SD_1 the standard deviation from study 1 and SD_2 is the standard deviation from study 2. M_1 is the mean from study 1 and M_2 is the mean from study 2.

$$\sqrt{\frac{\sum_{i=1}^{N} w_i\left(x_i-\bar{x}\right)^2}{\frac{(V-1)}{V}\sum_{i=1}^{N} w_i}} \tag{2}$$

Where N is the number of studies included in the group. w_i is a vector of the weighting for the study. x_i is the reported ICC value and $\bar{x}$ is the

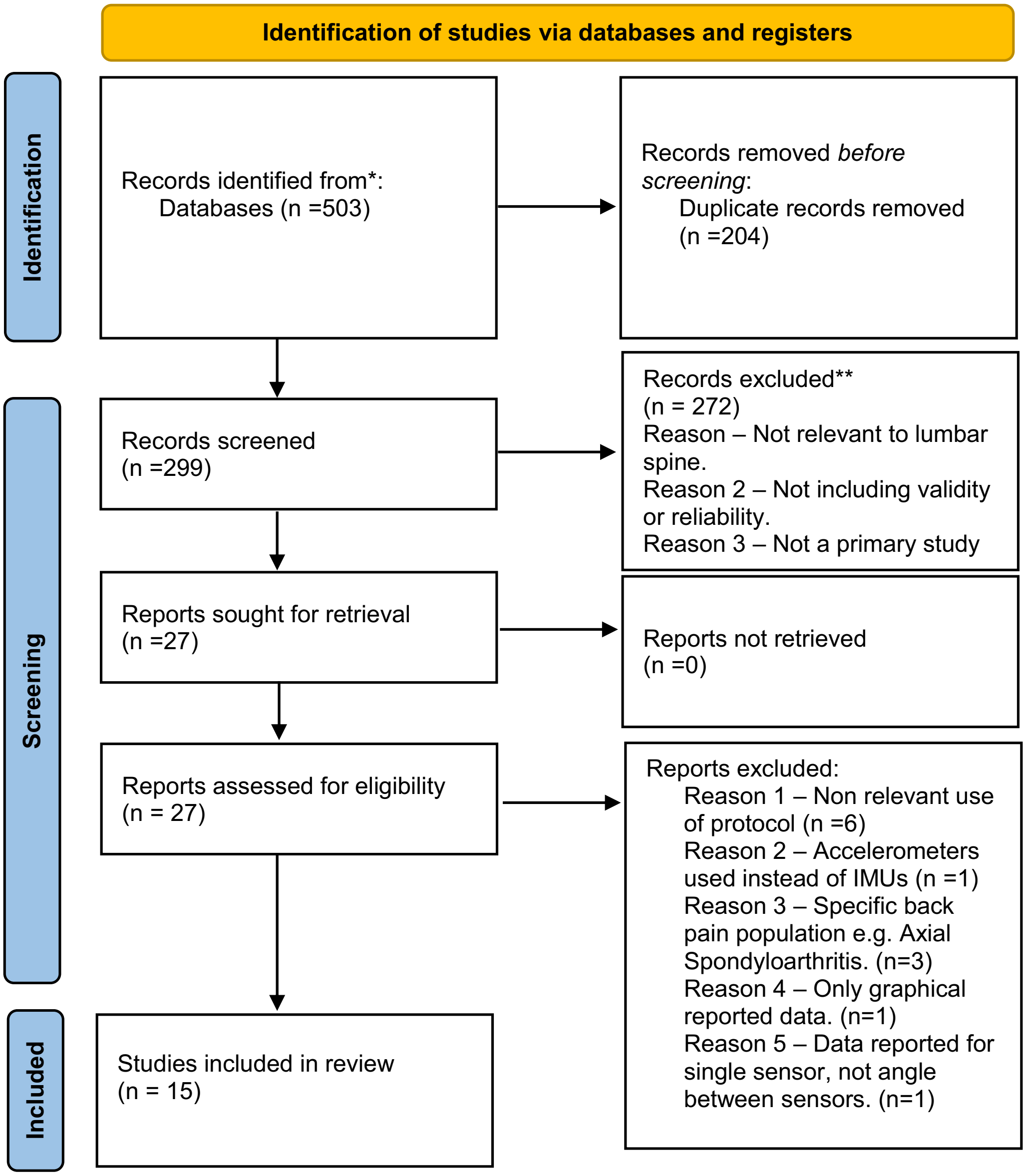

Fig. 1. PRISMA flow chart of the search strategy.

Table 2
Results of the QUADAS-2 appraisal tool for validity studies to determine risk of bias. *Y* = yes study passed with no bias, *N* = no study has risk of bias in this area, / = unclear.

Study	Unbiased selection of patients	Patients match review question	Unbiased interpretation of index test	Index test matches review question	Unbiased reference standard	Target condition matches review question	Unbiased patient flow
Wong 2007	Y	Y	Y	Y	Y	Y	Y
King 2009	/	Y	Y	Y	Y	Y	Y
Bauer 2015	Y	Y	Y	Y	Y	Y	Y
Mjosund 2017	Y	Y	/	Y	Y	Y	Y
Brice 2018	Y	Y	Y	Y	Y	Y	Y
Molnar 2018	/	Y	Y	Y	Y	Y	Y
Beange 2019	Y	Y	Y	Y	Y	Y	Y
Brice 2020	/	N	/	Y	Y	Y	Y
Senington 2020	/	Y	Y	Y	Y	Y	Y
Franco 2021	Y	Y	/	Y	Y	Y	Y
Brice 2022	Y	Y	Y	Y	Y	Y	Y

Table 3
Results of the QAREL appraisal tool for the reliability studies. *Y* = yes study has passed, *N* = no study has not passed, / = unclear, *N/A* = not applicable to study.

Study	Subjects representative	Raters representative	Raters blinded to other raters	Raters blinded to prior results	Raters blinded to reference standard	Raters blinded to clinical information	Raters blinded to additional cues	Order of examination varied	Time interval between measurements	Test applied correctly	Appropriate statistical methods
Williams 2013	Y	Y	N/A	N/A	N/A	N/A	/	N/A	Y	Y	Y
Bauer 2015	Y	Y	N/A	/	/	N/A	N	Y	Y	Y	Y
Yun 2015	Y	Y	/	/	N/A	N/A	/	/	Y	Y	Y
Bauer 2016	Y	Y	/	/	N/A	N/A	N	N/A	Y	Y	Y
Graham 2020	Y	Y	/	/	/	N/A	/	N/A	Y	Y	Y
Senington 2020	Y	Y	N/A	N/A	N/A	N/A	/	N/A	Y	Y	Y
Franco 2021	Y	Y	N/A	N/A	N/A	N/A	/	N/A	Y	Y	Y

Table 4A
Data extraction for the 16 included studies in this review, the movements and measurements included. Root mean square (RMSE), intraclass correlation (ICC), dependability, Pearson's correlation, and validity ICC.

Study	Design	Total sample size	Sample composition	Equipment	Sensor location	Comparator (Validity)	Testing procedure	Movements tested	Results RMSE (SD) ROM	ICC ROM	Dependability	Pearson's correlation	Validity ICC
Wong 2007	Observational validity	9	4 Females	3 Kionix sensor modules.	T1 T12 S2	Vicon motion capture.	Three trials completed for each movement of lumbar flexion and rotation.	Flexion extension				0.981	
								Lateral flexion				0.985	
			5 Males					Sit to stand				0.966	
King 2009	Observational validity	5	5 Males	3 BSN nodes	T12, sacrum, and middle of outer thigh	BTS bioengineering infra-red motion capture.	10 rowing strokes analysed, recording started after rowing had begun.	Rowing flexion	3.98°				
Williams 2013	Observational reliability	28	11 Females	2 3DM IMUs	L1 S1	–	Three trials, 3 kg box positioned on markers to allow exactly same position throughout lifting task, movement order was identical.	Flexion extension		0.99			
								Lateral flexion		0.98			
			17 Males					Rotation		0.94			
								Lifting		0.99			

IMU- Inertial measurement unit.
ICC- Intraclass corelation.
ROM- Range of motion.

Table 4B
Data extraction for the 16 included studies in this review, the movements and measurements included. Root mean square (RMSE), intraclass correlation (ICC), dependability, Pearson's correlation, and validity ICC.

Study	Design	Total sample size	Sample composition	Equipment	Sensor location	Comparator (Validity)	Testing procedure	Movements tested	Results RMSE (SD) ROM	ICC ROM	Dependability	Pearson's correlation	Validity ICC
Bauer 2015	Observational validity and reliability	46	24 Females	3 Valedo IMUs	L1 S2 and middle of outer thigh	Vicon motion capture	Randomized trial order per participant. Instructed to move as far as possible at their preferred speed.	Flexion extension	4.25° (2°)		0.72		
			22 Male					Lateral flexion	1.85° (0.65°)		0.9		
Yun 2015	Observational reliability	19	19 Males	2 MEMS-IMUs	T12 and sacrum.	–	Instruction not to fix pelvic motions, Practiced 6 lumbar movements. Verbal instruction given to keep each movement to a similar time length.	Flexion extension		0.978			
								Lateral flexion		0.965			
								Rotation		0.944			
Bauer 2016	Observational reliability	23	4 Females	4 Valedo IMUs	T1 L1 S2	–	Randomized trial order per participant, 30 min to perform all tests.	Flexion extension			0.86		
			19 Males					Lateral flexion			0.69		

IMU- Inertial measurement unit.
ICC- Intraclass corelation.
ROM- Range of motion.

Table 4C
Data extraction for the 16 included studies in this review, the movements and measurements included. Root mean square (RMSE), intraclass correlation (ICC), dependability, Pearson's correlation, and validity ICC.

Study	Design	Total sample size	Sample composition	Equipment	Sensor location	Comparator (Validity)	Testing procedure	Movements tested	Results RMSE (SD) ROM	ICC ROM	Dependability	Pearson's correlation	Validity ICC
Mjøsund 2017	Observational validity	34	16 Females	2 ViMove motion sensors	T12 and S2.	Vicon motion capture.	Movements performed to comfortable end of range and position held for 2 s.	Flexion extension	1.265° (0.67°)				
			18 Males					Lateral flexion	0.875° (0.47°)				
Brice 2018	Observational validity	5		3 IMeasureU IMUs	T3, PSIS, and middle of shank	Vicon motion capture	Maximal effort and pace of discus throw over 10 throws.	Discus Rotation	11.0°				
Molnar 2018	Observational validity	8		6 Musclelab IMUs	2 columns of 3 L1-L5	Vicon motion capture.	Between movements rest phase of at least seven seconds before performing maximum ROM.	Flexion extension	0.675° (0.2°)				
								Lateral flexion	0.85° (0.7°)				
								Rotation	1.25° (0.8°)				

IMU- Inertial measurement unit.
ICC- Intraclass corelation.
ROM- Range of motion.

Table 4D
Data extraction for the 16 included studies in this review, the movements and measurements included. Root mean square (RMSE), intraclass correlation (ICC), dependability, Pearson's correlation, and validity ICC.

Study	Design	Total sample size	Sample composition	Equipment	Sensor location	Comparator (Validity)	Testing procedure	Movements tested	Results RMSE (SD) ROM	ICC ROM	Dependability	Pearson's correlation	Validity ICC
Beange 2019	Observational validity	10	4 Females 6 Males	2 MetaMotionR IMUs.	T10/12 and S2	Vicon motion capture.	35 flexion extension cycles, touch two targets in sync with a metronome at 0.5 Hz. One target at shoulder height another at 50 cm anterior to the knee.	Flexion extension					Local Dynamic Stability- 0.807 Lumbopelvic Coordination-0.963
Brice 2020	Observational validity	17	9 Females 8 Males	5 IMeasureU IMUs	C7 T2 T7 pelvis and middle of sternum.	Vicon motion capture.	3 trials at a slow self-selected speed, then two maximal speed trials.	Dynamic Rotation	2.3° (0.6°)				

IMU- Inertial measurement unit.
ICC- Intraclass corelation.
ROM- Range of motion.

weighted mean of the reported ICC values. V is the number of non-zero weighted values.

3. Results

A total of 503 studies were identified through the systematic search of the literature, of which 15 studies met the inclusion criteria. Of these, eight reported specifically on validity [34–41], four on reliability [42–45], and three on both [46–48]. In total, 305 participants were included, of which 206 were male and 99 female, with an age range of 18–67 years old. Four out of 15 studies utilised individuals with LBP resulting in a total of 109 participants.

3.1. Validity

Nine studies reported validity and six were able to be integrated into the quantitative synthesis of RMSE [34–36,40,46,48]. Three studies were not included in the synthesis because they reported on different movements or reported results in measurements other than RMSE. Of the included studies, one reported validity data on participants with LBP [34] and five on healthy participants [35,36,40,46,48].

3.1.1. Flexion extension

The quantitative synthesis included four studies for the validity of flexion extension [34,40,46,48], (see Fig. 2a). One validity study included participants with LBP [34]. Regarding participants with LBP, the weighted mean RMSE was 1.3°, and for healthy participants was 2.9°. The weighted mean combining the studies resulted in an RMSE (95 %CI) of 2.4° (1.9–2.9°) where the average range of motion (ROM) was 71.2° yielding a percentage error of 3.3 %.

3.1.2. Lateral flexion

In the quantitative synthesis of four studies for validity [34,40,46,48] (see Fig. 2b), only one study investigated validity in individuals with LBP [34]. The weighted mean for RMSE of lateral flexion for individuals with LBP was 0.9° and for individuals without LBP was RMSE (95 %CI) 1.8° (1.6–2.0°). Overall, from all studies, the RMSE (95 %CI) was 1.5° (1.2–1.8°) where the average ROM was 43.1° yielding a percentage error of 3.5 %.

3.1.3. Rotation

The quantitative synthesis included two studies for validity [36,40] (see Fig. 2c), of which there were no reported results in participants with LBP. The weighted mean RMSE (95 %CI) was 2.0° (1.6–2.3°), where the average ROM was 15.2° producing a percentage error of 13.2 %.

3.1.4. Complex movements

Validity for complex movements (involving movements in all three planes) was reported in RMSE in four studies [35–38]. Two involved measuring lumbar movement in rowing [37,38], one in discus throwing [35], one investigating fast rotations [36]. The weighted mean for reported RMSE (95 % CI) for the complex movements combined was 2.0° (1.3–2.8°).

3.2. Reliability

Seven of the studies reported within session reliability and four were able to be included in the quantitative synthesis using ICC [42,43,45,48]. Three studies were not included in the synthesis because they reported on different movements, or reported results in measurements other than ICC. Of the included studies, two reported reliability data in participants with LBP [42,45] and two on healthy participants [43,48]. The ICC across the studies of participants with LBP ranged from 0.49 to 0.98 and for healthy participants ranged from 0.94 to 0.99, with three studies reflecting a near-perfect correlation [42,43,48].

Table 4E

Data extraction for the 16 included studies in this review, the movements and measurements included. Root mean square (RMSE), intraclass correlation (ICC), dependability, Pearson's correlation, and validity ICC.

Study	Design	Total sample size	Sample composition	Equipment	Sensor location	Comparator (Validity)	Testing procedure	Movements tested	Results RMSE (SD) ROM	ICC ROM	Dependability	Pearson's correlation	Validity ICC
Franco 2020	Observational validity and reliability	11	11 Males	5 Avanti IMUs	T1 T6 L1 S2 and forehead.	Qualisys motion capture	Each movement repeated 3 times, performed at a slow constant pace focusing on maximal ROM.	Flexion extension	1.65° (0.55°)	0.99			
								Lateral flexion	2.4° (1°)	0.99			
Graham 2020	Observational reliability	30	19 Females	2 Meylan IMUs	T8 S2	–	Movements completed to beat of metronome at 0.28,0.24 and 0.14 Hz respectively.	Flexion extension		0.49			
								Lateral flexion		0.71			
			11 Males					Rotation		0.88			
Senington 2021	Observational validity and reliability	40	40 Males	4 Thetametrics IMUs	T1 L1 S2 middle shank	Vicon motion capture.	6 maximal effort warm up bowls then 6 maximal effort recording bowls.	Lumbar flexion in bowling		0.93		0.99	
								Lumbar lateral flexion in bowling		0.64		0.95	
								Lumbar rotation in bowling		0.67		0.61	
Brice 2022	Observational validity	8	8 Females	4 IMeasureU IMUs	T1 T7 L2 S2	Vicon motion capture.	Participants rowed for 1 min at rates of 20,24,28, and 32 strokes per minute. Breaks at 18 strokes per minute between tests.	20 SPM	2.48° (0.54°)				
								24 SPM	2.84° (0.88°)				
			0 Males					28 SPM	2.31° (0.92°)				
								32 SPM	2.82° (0.93°)				

IMU- Inertial measurement unit.
ICC- Intraclass corelation.
ROM- Range of motion.

Table 5
Further reliability results from the studies.

Study	Total sample size	Population	Movement tested	Standard error measurement (°)	Minimum detectable difference (°)
Bauer 2016	23	LBP	Flexion extension		5.1
			Lateral flexion		4.25
Franco 2020	11	LBP	Flexion extension	0.55	
			Lateral flexion	1.1	
Graham 2020	30	LBP	Flexion extension	0.016	0.024
			Lateral flexion	0.007	0.011
			Rotation	0.012	0.018
Senington 2021	40	Healthy	Flexion extension	4.02	11.14
			Lateral flexion	4.46	12.36
			Rotation	4.82	13.36

LBP – Low back pain.

3.2.1. Flexion extension

The quantitative synthesis included four studies for reliability [42,43,45,48], (Fig. 2d). Regarding participants with LBP the weighted mean ICC (95 %CI) was 0.77 (0.71–0.83), and 0.99 (0.99–0.99) for healthy participants. The overall weighted mean for ICC (95 %CI) of flexion extension was 0.84 (0.72–0.95).

3.2.2. Lateral flexion

The quantitative synthesis included three studies for reliability [42,43,48], (Fig. 2e) with one study including individuals with LBP [42]. The weighted mean for reliability was 0.98 for participants with LBP and ICC (95 %CI) 0.96 (0.92–0.99) for participants without LBP, and 0.97 (0.97–0.98) when combined.

3.2.3. Rotation

The quantitative synthesis included three studies for reliability [42,43,45], (Fig. 2f). Two of the reliability studies were from LBP populations [42,45]. ICC (95 %CI) for LBP participants was 0.92 (0.92–0.93) and for healthy participants 0.94. Combined, the weighted mean (95 %CI) was 0.93 (0.90–0.95).

3.2.4. Complex movements

Reliability in complex movements was completed by two studies in this review, lumbar movement in cricket [47], and lifting movements [42]. The weighted mean ICC (95 % CI) value for these movements was 0.84 (0.81–0.88).

Orientation and therefore angles produced from the IMUs were calculated using different methods of sensor fusion of the accelerometer, gyroscope and magnetometer with five using a Kalman algorithm [49]: [36–39,43], three using a Madgwick algorithm [50,44–46], two used quaternion representation of orientation [35,48], two completed a fusion of just the accelerometer and gyroscope [40,41].

The location of the sensors ranged from T3-PSIS [35], to L1-L5 [43]. The most common placement of the two sensors over the lumbar spine was L1-Sacrum [42,44,46–48].

3.3. Methodological quality

A third of the studies that reported on measures of validity had the potential for bias around the selection of the sample used [36,38,40,47]. They used a purposeful sampling method rather than open sampling, therefore potentially limiting the generalisability of the findings to the wider population. However, as this review contains studies with both sampling methods, i.e., it includes studies in specific populations (cricket fast bowlers), as well as in individuals with and without LBP wider generalisability could be possible. Rater representativeness was assessed in studies where the raters and/or their professions were described clearly, or if an ICC (2,1) was completed, as the raters for this version of ICC must be representative of the population. If the profession stated physiotherapist, or clinician, this was agreed to be representative of the target population.

The main potential threat of bias shared across reliability studies relates to the use of blinding. This was evident in either raters or assessors being unblinded to reference values and prior results. Blinding of raters and assessors is important to prevent attempts to align current values to previous values or repeat tests when undesirable values are produced. However, with the optoelectronic and IMU systems such methods are automated and post-processed, therefore limiting the influence of this bias across the results.

4. Discussion

This review, for the first time, presents a synthesis of reliability and validity estimates, along with a scope of the literature pertaining to the measurement of lumbar motion using IMUs. Furthermore, a critical strength of the review is that it includes data for individuals with non-specific LBP, enabling researchers and clinicians to understand these metrics in a specific and common target population. Overall, the results demonstrate that IMUs can be considered valid and reliable for measuring lumbar ROM, in individuals with and without LBP. This was observed regardless of the plane of movement, demonstrating the applicability to use IMUs in clinical assessments. Furthermore, the results remain applicable to more complex multidimensional (multi-plane) movements, such as those seen in sports or daily functional tasks. This demonstrates the ability of the technology to assess movement in dynamic environments to an excellent standard, which is important when exploring mechanisms underpinning of LBP.

Despite this overarching synthesis statement, some nuances around the findings and application of this technology warrant further discussion. The method of measuring validity and reliability differed across the studies. Some studies reported RMSE values as a measure of the error between the two systems, whilst others opted to explore the relationship (correlation) between the two systems, demonstrating excellent correlations [41,47]. Such bivariate correlations are only able to explore the relationship of one variable to another, and not how similar or different the values are. An excellent correlation is possible even in the presence of large differences between the two systems. RMSE estimates overcome this limitation by providing greater clarity over how similar or different specific values for each system are, however again by design, do not provide estimates of the relationship between the values. The results from this review, through the cumulative synthesis of findings from across the studies demonstrate that IMUs offer a measure that is both highly correlated to the optoelectronic gold standard but also highly similar based on RMSE values. Therefore, regardless of the method, good-excellent results for validity were demonstrated.

Reliability was approached variably across the studies with most of the studies reporting ICC values [42,43,45,47,48]. Two studies used an alternate approach to reliability estimation, utilising the index of dependability, which enabled the authors to model the effect of manipulating the number of trials or repeated movements [44,46]. This has the advantage of removing unnecessary testing, especially when individuals may be experiencing pain, and these studies suggest one trial may be enough due to high-reliability coefficients 0.69–0.9 [44,46]. Therefore, regardless of the method employed to estimate reliability, good-excellent results were determined. Further reliability data (Table 5) show SEMs between 0.01–4.82°, with the two studies on LBP populations producing the lower error values compared to the healthy population. This may be due to a smaller ROM of LBP participants,

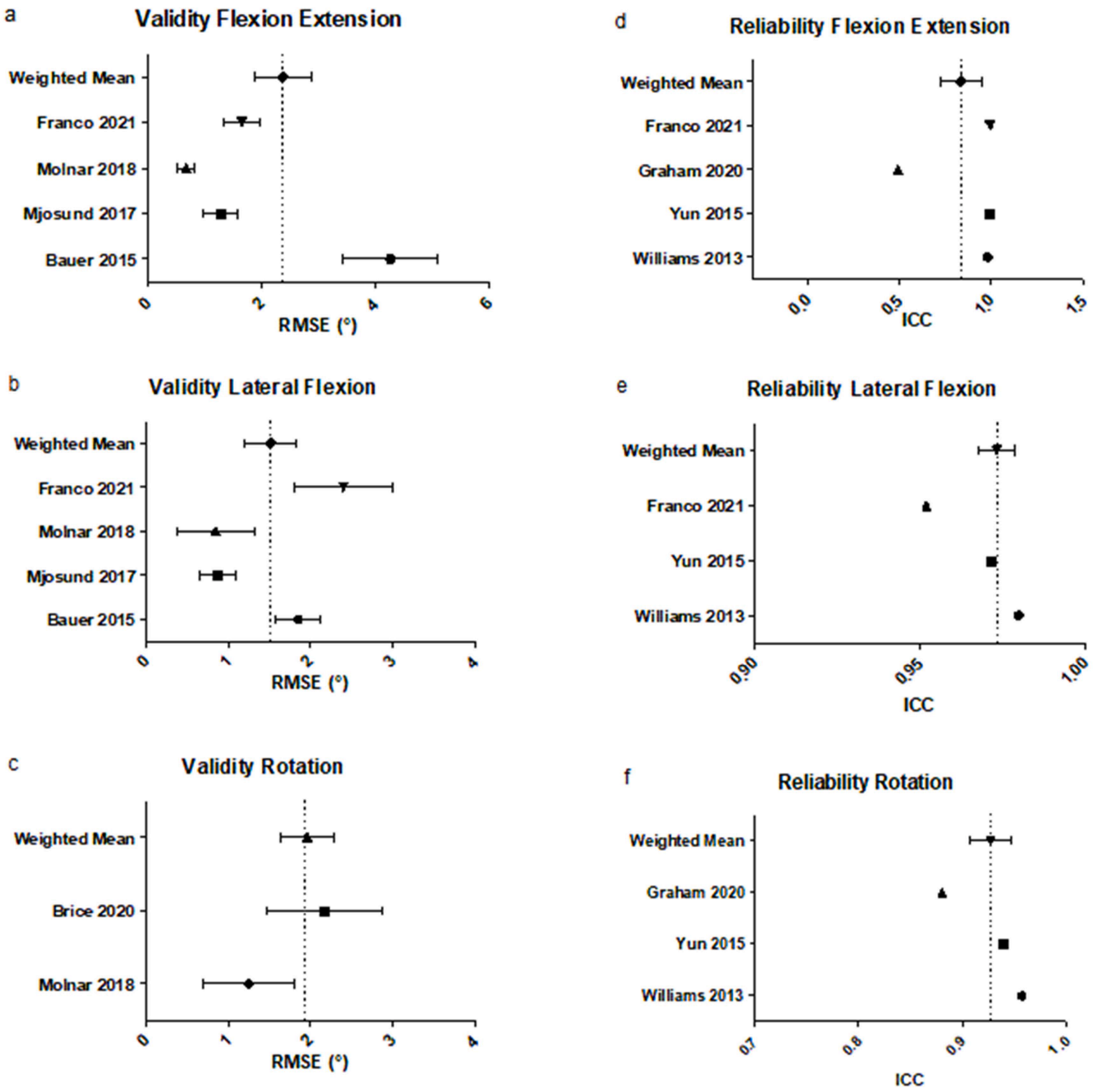

Fig. 2. Forest plots of the weighted means and confidence intervals from the quantitative synthesis. a- Validity of flexion extension, b- Validity of lateral flexion, c- Validity of rotation, d- Reliability of flexion extension, e- Reliability of lateral flexion, f- Reliability of rotation.

however the results show that the error values are low compared to the range of motion, further contributing to the reliability of the IMUs to measure the lumbar spine.

Approaches to validity and reliability used in the articles included in this review had a strong emphasis on a single value or point in time, (e.g., peak ROM). For example, RMSE estimates were for the 'peak' ROM of the optoelectronic system compared to IMUs. Similarly, ICC values were commonly determined from peak ROM values also. Such analysis does not consider the similarity of the data across time. It is entirely possible to observe very different movement behaviour which converges at the peak, yielding good estimates of both validity and reliability. To understand the similarity in the ROM across time, cross-correlation (rather than ICC) with related RMSE would serve to better estimate relationships and differences in ROM across time between the optoelectronic and IMU systems, representing validity for the whole movement time curve [51]. Similarly, the coefficient of multiple correlation with related RMSE, would explore and quantify the similarities and differences between repeated movement time curves and represent an estimate of reliability [42]. As IMUs have the potential to explore the movement of the lumbar spine during normal daily function and as normal daily function does not often reach peak ROM [52], determining if validity and reliability estimates remain applicable at other ranges would be beneficial.

The quantitative synthesis of RMSE for validity demonstrates $<2.4^\circ$ on average, with rotation offering slightly higher relative error. As a percentage, the largest difference is seen with rotation which seems related to the small ROM in this plane rather than a large error observed. However, this does result in larger relative errors for this motion. Sensor error relates to the internal consistency of the device during repeated known measurements and is often reported by IMU manufacturers to range from 0.5 to 2°. Often data specification sheets report measures of Yaw/Heading to be less accurate than those of pitch and roll, presumably related to the lack of contribution from the accelerometers to the fusion, which may explain some of the error here [48,53].

The error related to the human-sensor interface relates specifically to the attachment of the device, where relative movement between the skin and sensor results in error. A variety of attachment methods were observed in the studies, most commonly either attaching sensors directly to the skin with double-sided tape [36,42], or via a plastic plate to increase the attachment footprint [39]. Importantly some articles chose to 'piggyback' one measurement system on another (e.g., 3D motion capture markers on an IMU), thereby minimising the influence of this interface error, as changes in the single attachment affect both systems. There was no discernible difference in values for this method compared to non-piggy-backed methods suggesting either attachment method is both valid and reliable. The differences in the placement of the two IMUs in each study to measure across the lumbar spine has the potential to create differences between the recorded values and was variable in the studies included in this review. However, there was no distinguishable difference between the locations therefore, the placement of the two IMUs does not alter the validity or reliability. However, for accurate measurement of the lumbar spine alone, with minimal interference from the thoracic spine, it would be recommended to place the IMUs on the skin over the anatomical limits of the lumbar spine (L1 and Sacrum), which was the most common setup in the studies of this review.

Another potential source of error is human error which represents the variability in humans repeating movements. This natural variance is important in the measures of reliability but less in methods that simultaneously measure the two systems. Evidence from these studies suggests most repeated movements are highly similar, except for flexion and extension in individuals with LBP. Good reliability estimates were calculated for individuals with LBP, but these were excellent in individuals without LBP. This may be representative of movements where there is the expectation of pain or actual provocation of pain, as this can result in altered movement behaviour [24].

IMUs offer several advantages over optoelectronic measurement systems. They are cheaper than a motion capture system, with single sensors retailing from £100 to £1000. They do not require line-of-sight, which offers benefits for applications where the low back is not exposed. Sensors are quick to attach and require little calibration making them an attractive option for clinical and research applications. Two studies used fusion of the accelerometer and gyroscope only [40,41]. There were no differences in the results of this study compared to the others, demonstrating the potential for using IMUs in multiple environments, for example in clinical areas that may be compromised by magnetic interference. Literature demonstrates development of protocols that estimate joint angle from IMU readings without the magnetometer, building on the idea that IMUs can be used in the future regardless of the setting [54,55]. Now that validity and reliability have been synthesised users can have confidence that errors are well understood compared to the existing gold standard. However, there are several limitations of IMUs as compared to optoelectronic measurements, most notably the absence of directly measured location, providing solely orientation. Computationally, location is derivable however the reliability and validity of this is beyond the scope of this review.

This study has shown that IMUs have the potential to measure through time, and from this review it is known that they can correctly measure the lumbar spine. Furthermore, the values reported in this study can be used by clinicians and researchers. Understanding the error estimates enables clinicians to analyse movement and be confident of true change compared to sensor error. For researchers such values can be used in effect size calculations and confidence estimates, giving an important contribution to future design of studies involving IMUs and individuals with LBP.

Future studies should look to understand through movement measurements of the lumbar spine further using IMUs. This will allow greater understanding of movement patterns and load management. With the accessibility and versatility of IMUs, research into LBP can be completed in ways and locations where quantification was not previously possible. Multiple IMUs may be used to understand health conditions, or even sporting performance, injury risk and movement efficiency. Research needs to be completed to understand the effect of speed on IMU joint angle measurement.

5. Strengths and limitations

This review has made several contributions to the literature. Firstly, it has provided, through quantitative synthesis, estimates of error for using IMUs to measure lumbar movements as compared to the gold standard. Researchers and clinicians can utilise such values to determine if such devices meet the needs of their application. Secondly, this study has provided estimates of reliability, demonstrating that in individuals with and without LBP, that such devices can reliably measure lumbar movement. Another strength of this paper is the use of a weighted synthesis, which places a greater 'weight' on the values from studies with larger sample sizes. As such the proportion of each study's sample size to the total cumulative sample size is considered as part of the calculation. Without such a method a study with a sample size of 5 would be treated with equal 'weight' as one with a sample size of 100. For example, in this review the RMSE for flexion extension in healthy participants as an arithmetic mean is 2.2° compared to the used weighted mean of 2.9°. This arithmetic mean would result in erroneous conclusions, in this case suggesting a reduction in error, whereas the weighted mean reflects a better representation of true error.

Finally, these findings seem to hold true for different sensors and different attachments. One of the limitations of this review is that not all studies reported the same movements or the same units. This meant that some studies were not included in the synthesis and as such the weighted mean was limited to fewer studies. As no funds were available for English language translation, only studies in this language were used to avoid misinterpretation of other studies, which may have limited the number of studies used in this review.

6. Conclusion

The aim of this review was to establish concurrent validity (compared to the gold standard) and the repeated measures reliability of IMUs for measurement of the movement of the lumbar spine in individuals with and without LBP. The results demonstrated excellent/good validity against the gold standard for all single plane and complex (multi-plane) movements, in individuals with and without LBP. The most common validation method, mirroring clinical practice, uses 'peak' ROM, yet 'peak' ROM is seldom seen in everyday activities. Instead, validity for the whole movement time curve would allow a better estimate of movement over time and allow the progression of metrics for lumbar spine assessment and rehabilitation. IMUs offer clinicians and researchers a valid and reliable measurement of motion in the lumbar spine compared to the current laboratory method.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical approval

Not required.

CRediT authorship contribution statement

Frederick A. McClintock: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Andrew J. Callaway:** Conceptualization, Methodology, Supervision, Formal analysis, Writing – review & editing. **Carol J. Clark:** Conceptualization,

Methodology, Supervision, Formal analysis, Writing – review & editing. **Jonathan M. Williams:** Conceptualization, Methodology, Supervision, Formal analysis, Writing – review & editing.

Declaration of competing interest

The authors declare no competing interest. All the authors have no affiliations with any organisations discussed in this review.

References

[1] Vos T, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388(10053):1545–602. https://doi.org/10.1016/S0140-6736(16)31678-6.

[2] Wood L, Hendrick PA. A systematic review and meta-analysis of pain neuroscience education for chronic low back pain: short-and long-term outcomes of pain and disability. EJP 2019;23(2):234–49. https://doi.org/10.1002/ejp.1314.

[3] Hoy D, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. ARD 2014;73(6):968. https://doi.org/10.1136/annrheumdis-2013-204428.

[4] Ganesan S, et al. Prevalence and risk factors for low back pain in 1,355 young adults: a cross-sectional study. ASJ 2017;11(4):610–7. https://doi.org/10.4184/asj.2017.11.4.610.

[5] Itz CJ, et al. Clinical course of non-specific low back pain: a systematic review of prospective cohort studies set in primary care. EJP 2013;17(1):5–15. https://doi.org/10.1002/j.1532-2149.2012.00170.x.

[6] Urits I, et al. Low back pain, a comprehensive review: pathophysiology, diagnosis, and treatment. Curr Pain Headache Rep 2019;23(3):1–10. https://doi.org/10.1007/s11916-019-0757-1.

[7] Luomajoki H, et al. Movement control tests of the low back; evaluation of the difference between patients with low back pain and healthy controls. BMC MD 2008;9(1):170. https://doi.org/10.1186/1471-2474-9-170.

[8] Morris M, Fau - Pellow J, et al. Physiotherapy and a homeopathic complex for chronic low-back pain due to osteoarthritis: a randomized, controlled pilot study. Altern Ther Health Med 2016;(1078–6791).

[9] Garcia AN, et al. Effectiveness of back school versus McKenzie exercises in patients with chronic nonspecific low back pain: a randomized controlled trial. PTJ 2013;(1538–6724). https://doi.org/10.2522/ptj.20120414.

[10] Cuesta-Vargas AI, Galán-Mercant A, Williams JM. The use of inertial sensors system for human motion analysis. PTR 2010;15(6):462–73. https://doi.org/10.1179/1743288×11Y.0000000006.

[11] Matthew RP, Seko S, Bajcsy R, Lotz J. Kinematic and kinetic validation of an improved depth camera motion assessment system using rigid bodies. IEEE JBHI 2018;23(4):1784–93.

[12] Topley M, Richards JG. A comparison of currently available optoelectronic motion capture systems. J Biomech 2020;106:109820. https://doi.org/10.1016/j.jbiomech.2020.109820.

[13] Weygers I, et al. Inertial sensor-based lower limb joint kinematics: a methodological systematic review. Sensors 2020;20(3):673. https://doi.org/10.3390/s20030673.

[14] van Dijk MJ, Smorenburg NT, Heerkens YF, Mollema J, Kiers H, Nijhuis-van der Sanden MW, Visser B. Assessment instruments of movement quality in patients with non-specific low back pain: a systematic review and selection of instruments. Gait Posture 2020;76:346–57. https://doi.org/10.1016/j.gaitpost.2019.12.027.

[15] Bolink SAAN, et al. Validity of an inertial measurement unit to assess pelvic orientation angles during gait, sit-stand transfers and step-up transfers: comparison with an optoelectronic motion capture system. Med Eng Phys 2016;38(3):225–31. https://doi.org/10.1016/j.medengphy.2015.11.009.

[16] Taylor L, Miller E, Kaufman KR. Static and dynamic validation of inertial measurement units. Gait Posture 2017;57:80–4. https://doi.org/10.1016/j.gaitpost.2017.05.026.

[17] Höflinger F, et al. A wireless micro inertial measurement unit (IMU). IEEE TIM 2013;62(9):2583–95. https://doi.org/10.1109/TIM.2013.2255977.

[18] Palermo E, Rossi S, Marini F, Patanè F, Cappa P. Experimental evaluation of accuracy and repeatability of a novel body-to-sensor calibration procedure for inertial sensor-based gait analysis. Measurement 2014;52:145–55. https://doi.org/10.1016/j.measurement.2014.03.004.

[19] Šlajpah S, Kamnik R, Munih M. Kinematics based sensory fusion for wearable motion assessment in human walking. Comput Methods Programs Biomed 2014;116(2):131–44. https://doi.org/10.1016/j.cmpb.2013.11.012.

[20] Slade P, Habib A, Hicks JL, Delp SL. An open-source and wearable system for measuring 3D human motion in real-time. IEEE TBE 2021;69(2):678–88. https://doi.org/10.1101/2021.03.24.436725.

[21] Novy DM, Simmonds MJ, Olson SL, Lee CE, Jones SC. Physical performance: differences in men and women with and without low back pain. Arch Phys Med Rehab 1999;80(2):195–8.

[22] Marras WS, Lewis KE, Ferguson SA, Parnianpour M. Impairment magnification during dynamic trunk motions. Spine 2000;25(5):587–95.

[23] Shum GL, Crosbie J, Lee RY. Three-dimensional kinetics of the lumbar spine and hips in low back pain patients during sit-to-stand and stand-to-sit. Spine 2007;32(7):E211–9.

[24] Williams JM, Haq I, Lee RY. An experimental study investigating the effect of pain relief from oral analgesia on lumbar range of motion, velocity, acceleration and movement irregularity. BMC MD 2014;15(1):1–9.

[25] Lebel K, Boissy P, Hamel M, Duval C. Inertial measures of motion for clinical biomechanics: comparative assessment of accuracy under controlled conditions-effect of velocity. PLoS ONE 2013;8(11):e79945.

[26] Cottam DS, Campbell AC, Davey MPC, Kent P, Elliott BC, Alderson JA. Measurement of uni-planar and sport specific trunk motion using magneto-inertial measurement units: the concurrent validity of Noraxon and Xsens systems relative to a retro-reflective system. Gait Posture 2022;92:129–34.

[27] Kobsar D, Charlton JM, Tse CT, Esculier JF, Graffos A, Krowchuk NM, Thatcher D, Hunt MA. Validity and reliability of wearable inertial sensors in healthy adult walking: a systematic review and meta-analysis. JNER 2020;17:1–21.

[28] Horsley BJ, et al. Does site matter? Impact of inertial measurement unit placement on the validity and reliability of stride variables during running: a systematic review and meta-analysis. Sports Med 2021;51(7):1449–89. https://doi.org/10.1007/s40279-021-01443-8.

[29] Papi E, Koh WS, McGregor AH. Wearable technology for spine movement assessment: a systematic review. J Biomech 2017;64:186–97. https://doi.org/10.1016/j.jbiomech.2017.09.037.

[30] Page MJ, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021;10(1):1–11. https://doi.org/10.1136/bmj.n71.

[31] Quadas 2Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, QUADAS-2 Group*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. AIM 2011;155(8):529–36.

[32] Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). JCE 2010;63(8):854–61.

[33] Cochrane. 6.5.2.10 combining groups. 2022 14 October 2022]; Available from: https://training.cochrane.org/handbook/current/chapter-06#section-6-5.

[34] Mjøsund HL, et al. Clinically acceptable agreement between the ViMove wireless motion sensor system and the Vicon motion capture system when measuring lumbar region inclination motion in the sagittal and coronal planes. BMC MD 2017;18(1):124. https://doi.org/10.1186/s12891-017-1489-1.

[35] Brice SM, Hurley M, Phillips EJ. Use of inertial measurement units for measuring torso and pelvis orientation, and shoulder–pelvis separation angle in the discus throw. Int J Sports Sci Coach 2018;13(6):985–92. https://doi.org/10.1177/1747954118778664.

[36] Brice SM, et al. Comparing inertial measurement units and marker-based biomechanical models during dynamic rotation of the torso. EJSS 2020;20(6):767–75. https://doi.org/10.1080/17461391.2019.1666167.

[37] Brice SM, Millett EL, Philippa B. The validity of using inertial measurement units to monitor the torso and pelvis sagittal plane motion of elite rowers. J Sports Sci 2022:1–9. https://doi.org/10.1080/02640414.2022.2042146.

[38] King RC, et al. Body sensor networks for monitoring rowing technique. IEEE 2009:251–5. https://doi.org/10.1109/BSN.2009.60.

[39] Beange KHE, et al. Concurrent validity of a wearable IMU for objective assessments of functional movement quality and control of the lumbar spine. J Biomech 2019;97:109356. https://doi.org/10.1016/j.jbiomech.2019.109356.

[40] Molnar M, et al. A method for lower back motion assessment using wearable 6D inertial sensors. In: 2018 21st international conference on information fusion (Fusion); 2018. p. 799–806. https://doi.org/10.23919/ICIF.2018.8455828.

[41] Wong WY, Wong MS. Trunk posture monitoring with inertial sensors. Eur Spine J 2008;17(5):743–53. https://doi.org/10.1007/s00586-008-0586-0.

[42] Williams JM, Haq I, Lee RY. A novel approach to the clinical evaluation of differential kinematics of the lumbar spine. Man Ther 2013;18(2):130–5. https://doi.org/10.1016/j.math.2012.08.003.

[43] Yun W-S, et al. Individual characteristics of reliable lumbar coupling motions. Eur Spine J 2015;24(9):1917–25. https://doi.org/10.1007/s00586-015-4081-0.

[44] Bauer CM, et al. Reliability of lumbar movement dysfunction tests for chronic low back pain patients. Man Ther 2016;24:81–4. https://doi.org/10.1016/j.math.2016.02.013.

[45] Graham RB, Dupeyron A, van Dieën JH. Between-day reliability of IMU-derived spine control metrics in patients with low back pain. J Biomech 2020;113:110080. https://doi.org/10.1016/j.jbiomech.2020.110080.

[46] Bauer CM, et al. Concurrent validity and reliability of a novel wireless inertial measurement system to assess trunk movement. JEK 2015;25(5):782–90. https://doi.org/10.1016/j.jelekin.2015.06.001.

[47] Senington B, Lee RY, Williams JM. Validity and reliability of innovative field measurements of tibial accelerations and spinal kinematics during cricket fast bowling. Med Biol Eng Comput 2021;59:1475–84.

[48] Franco L, et al. A novel IMU-based clinical assessment protocol for axial spondyloarthritis: a protocol validation study. PeerJ 2021;9:e10623. https://doi.org/10.7717/peerj.10623.

[49] Kalman RE. A new approach to linear filtering and prediction problems. J Basic Eng 1960;82(1):35–45. https://doi.org/10.1115/1.3662552.

[50] Madgwick SO, Harrison AJ, Vaidyanathan R. Estimation of IMU and MARG orientation using a gradient descent algorithm. In: 2011 IEEE international conference on rehabilitation robotics. IEEE; 2011. p. 1–7.

[51] Williams JM, Haq I, Lee RY. Dynamic measurement of lumbar curvature using fibre-optic sensors. Med Eng Phys 2010;32(9):1043–9. https://doi.org/10.1016/j.medengphy.2010.07.005.

[52] Bible JE, et al. Normal functional range of motion of the lumbar spine during 15 activities of daily living. Clin Spine Surg 2010;23(2):106–12. https://doi.org/10.1097/BSD.0b013e3181981823.

[53] Vitali RV, McGinnis RS, Perkins NC. Robust error-state Kalman filter for estimating IMU orientation. IEEE Sensors 2020;21(3):3561–9. https://doi.org/10.1109/JSEN.2020.3026895.

[54] Kok M, Hol JD, Schön TB. An optimization-based approach to human body motion capture using inertial sensors. IFAC Proc Vol 2014;47(3):79–85. https://doi.org/10.3182/20140824-6-ZA-1003.02252.

[55] Rapp E, Shin S, Thomsen W, Ferber R, Halilaj E. Estimation of kinematics from inertial measurement units using a combined deep learning and optimization framework. J Biomech 2021;116:110229. https://doi.org/10.1016/j.jbiomech.2021.110229.

Frederick A. McClintock is studying to achieve a PhD from Bournemouth University, with a MSc Sports Biomechanics (Loughborough University), and a BSc Biological Sciences (Bournemouth University).

Andrew J. Callaway is a principal academic in Biomechanics and Performance Analysis at Bournemouth University, with a PhD in Biomechanics and Sports Technology (Bournemouth University).

Carol Clark is a Professor in Physiotherapy at Bournemouth University, with a Masters in Advanced Musculoskeletal Physiotherapy (University College London) and a PhD in Health Care (Bournemouth University).

Jonathan M. Williams is an Associate Professor at Bournemouth University, with a BSc (Hons) Physiotherapy (University of Southampton), a Master of Manipulative Therapy (Curtin University) and a PhD in Biomechanics (University of Roehampton).