RHEUMATOLOGY

Rheumatology 2020;59:3415–3423
doi:10.1093/rheumatology/keaa122
Advance Access publication 28 April 2020

Original article

# Validity and reliability of a sensor-based electronic spinal mobility index for axial spondyloarthritis

**Philip V. Gardiner [1], Dawn Small[1], Karla Muñoz-Esquivel[2], Joan Condell[2], Antonio Cuesta-Vargas[3], Jonathan Williams[4], Pedro M. Machado [5,6,7] and Juan L. Garrido-Castro [8]**

## Abstract

**Objective.** To evaluate the validity and reliability of inertial measurement unit (IMU) sensors in the assessment of spinal mobility in axial spondyloarthritis (axSpA).

**Methods.** A repeated measures study design involving 40 participants with axSpA was used. Pairs of IMU sensors were used to measure the maximum range of movement at the cervical (Cx) and lumbar (Lu) spine. A composite IMU score was defined by combining the IMU measures. Conventional metrology and physical function assessment were performed. Validation was assessed considering the agreement of IMU measures with conventional metrology and correlation with physical function. Reliability was assessed using intra-class correlation coefficients (ICCs).

**Results.** The composite IMU score correlated closely ($r = 0.88$) with the BASMI. Conventional Cx rotation and lateral flexion tests correlated closely with IMU equivalents ($r = 0.85$, 0.84). All IMU movement tests correlated strongly with BASFI, while this was true for only some of the BASMI tests. The reliability of both conventional and IMU tests (except for chest expansion) ranged from good to excellent. Test–retest ICCs for individual conventional tests varied between 0.57 and 0.91, in comparison to a range from 0.74 to 0.98 for each of the IMU tests. Each of the composite regional IMU scores had excellent test–retest reliability (ICCs=0.94–0.97), comparable to the reliability of the BASMI (ICC=0.96).

**Conclusion.** Cx and Lu spinal mobility measured using wearable IMU sensors is a valid and reliable assessment in multiple planes (including rotation), in patients with a wide range of axSpA severity.

**Key words:** reliability, spinal mobility, axial spondyloarthritis, sensor, inertial measurement unit

**Rheumatology key messages**

- Wearable IMU sensors show excellent reliability in the measurement of spinal mobility in axSpA patients.
- A composite 'IMU based metrology index, the IMU-ASMI', shows excellent reliability as an outcome score for axSpA spinal mobility.

## Introduction

It is widely recognized that spinal mobility should be measured as an outcome measure in axial spondyloarthritis (axSpA). The Assessment in SpondyloArthritis international Society (ASAS) has recommended spinal mobility as a core domain in both clinical practice and trials [1]. The European Medicines Agency stated that 'spinal mobility is of great importance in ankylosing spondylitis (AS) and constitutes the most specific domain because other domains are common with many other rheumatic diseases. Although it may be difficult to detect changes in spinal mobility on the short term, spinal mobility is considered an important measure to assess efficacy' [2]. The most frequently used spinal mobility tool is the BASMI [3]. This index is based on a

[1]Department of Rheumatology, Western Health and Social Care Trust, Londonderry, [2]School of Computing, Engineering and Intelligent Systems, Intelligent Systems Research Centre, Ulster University, Derry/Londonderry, UK, [3]Department of Physiotherapy, University of Malaga, IBIMA, Malaga, Spain, [4]Department of Human Sciences and Public Health, Bournemouth University, Bournemouth, UK, [5]Centre for Rheumatology and MRC Centre for Neuromuscular Diseases, University College London, [6]Department of Rheumatology, University College London Hospitals NHS Foundation Trust, [7]Department of Rheumatology, London North West University Healthcare NHS Trust, Northwick Park Hospital, London, UK and [8]Motion Analysis Laboratory, Maimonides Biomedical Research Institute of Cordoba (IMIBIC), Cordoba, Spain

Submitted 7 November 2019; accepted 20 February 2020

Correspondence to: Philip V. Gardiner, Department of Rheumatology, Western Health and Social Care Trust, Glenshane Road, Londonderry, BT47 6SB, UK. E-mail: philip.gardiner@westerntrust.hscni.net





CLINICAL SCIENCE

mixture of tests carried out using a tape measure and goniometer: only three of the five tests are tests of spinal mobility. The only movement test measured in degrees is cervical (Cx) rotation. Critics have highlighted floor effects with components of the BASMI [4], while others have highlighted its poor responsiveness to change and its dubious content validity [5–7]. A recent attempt to develop and validate another manual metrology tool (the Edmonton AS Metrology Index) showed some improvements over BASMI but responsiveness to change was still relatively poor [8]. Some researchers prefer to report the individual components of spinal mobility rather than the composite BASMI, but there is no consistent evidence that any one component is more responsive to change than the overall score.

Motion capture methods are widely regarded as the gold standard for the accurate and automated measurement of movement [9–12]. In 2004, Jordan *et al.* [13] used an electromagnetic measurement system (Fastrak) to measure range of movement in the shoulder and Cx spine in axSpA. A high level of reliability was demonstrated, especially in the Cx spine; however, such technology is known to suffer from metallic interference [14]. Garrido-Castro *et al.* [15] subsequently developed and validated the UCOTrack motion capture system to measure spinal mobility in axSpA. A spinal mobility score based on this system (the University of Cordoba AS Metrology Index—UCOASMI) has superior reliability and responsiveness in axSpA in comparison to the BASMI [16, 17]. However, this movement laboratory-based method is expensive and requires dedicated facilities and expertise to set up and to perform the tests.

Although the above methods may have little relevance to clinicians, novel inertial measurement unit (IMU) sensor technology promises to provide the clinician with advanced tools that are affordable, accurate and easy to use. Wearable devices incorporating these sensors should represent a significant step forward in the accurate measurement of spinal mobility. Current measurements based on the use of goniometers and tape measures are open to observer variability. Spinal mobility measures based on the use of tape measures do not directly measure the angle of movement and are therefore subject to variation between subjects due to anthropomorphic features such as height and leg length. These measures lack content validity as they cannot record potentially important aspects of spinal mobility such as spinal rotation [5–7]. Unlike traditional tools, IMU sensors can also be used to measure dynamic movement; that is, continuous variation of angles, the speed of movement as well as the maximal range of movement. Besides this, they can be used in the home or work environment. Early IMU devices were subject to errors but the use of combined sensors, filtering of 'noise' and compensation for drift gyroscope error enable accurate measurements as confirmed in tests against gold standard motion capture methods.

There is a growing body of evidence that IMU-based sensors can accurately measure spinal movement in normal individuals and those with back pain [18–21]. Ronchi tested a set of IMU sensors positioned according to the limits of the Modified-Modified Schöber Test and demonstrated excellent reliability in normal subjects, superior to the traditional tape measure method and to dual inclinometers [22]. The ViMove IMU system was based on that work but evolved further with the addition of lumbar (Lu) rotation and Cx movement tests to the protocol. Our choice of ViMove IMU sensors was based on strong validation studies in normal individuals and patients with back pain. These studies used a clearly defined method and careful placement of sensors across the Lu spine that seemed to parallel Schober's test, features that we felt would reduce variability. Furthermore, these sensors have been validated against a motion capture system [23], are approved for use in patients with back pain and the software is straightforward for the non-expert user. The primary objective of this study was to investigate the validity and reliability of an IMU sensor-based test of spinal movement in people with axSpA.

## Methods

People with axSpA were involved in the design and analysis of the study: discussions were held before the study protocol was finalized and the results have been shared with our patient research forum. The study was approved by the regional ethics committee (Office for Research Ethics Committees Northern Ireland) and was carried out in compliance with the Helsinki Declaration. It was registered with clinicaltrials.gov (NCT03159767). All participants gave informed consent to take part in the study. Clinical physiotherapists, with at least 2 years of experience in measuring axSpA patients, carried out clinical and sensor movement tests.

### Study sample

Participants over the age of 18 years with axSpA who fulfilled the ASAS classification criteria were included in the study. The selection was performed through 'convenience' sampling at clinics or physiotherapy sessions. Those with a history of spinal/hip surgery and those with a history of spinal fracture or a major scoliosis deformity were excluded. Severe joint or spinal pain at the time of the study resulted in exclusion. Information on age, sex, diagnosis, duration of disease and therapy was collected. The BASDAI, the BASG score and the BASFI questionnaires were completed. BASMI and chest expansion testing was carried out according to the ASAS handbook guidelines using a tape measure and goniometer [7]. The linear version of BASMI ($BASMI_{lin}$) was used and the values of each component recorded [24].

### Study design/procedures

A flow diagram of assessment is shown in Fig. 1. On the first visit (day 1), each participant had conventional metrology and sensor testing carried out three times. One

physiotherapist (Rater A) carried out a twin set of measurements an h apart. The sensors and any marks were removed between assessments and before re-application. The second physiotherapist (Rater B) – working in another room – carried out a third set of measurements without knowledge of previous results.

Participants were asked to return 1–2 weeks later at the same time of day for repeat conventional metrology and sensor testing (day 2).

### IMU sensor movement test procedures

The ViMove wireless sensor kit (DorsaVi, Australia) is a wireless IMU system comprised of two wireless movement sensors each containing a gyroscope, a magnetometer and an accelerometer (Fig. 2). These were paired with a pocket wireless device recording at a rate of 20 Hz and connected to a laptop, so that the angular displacement of each sensor could be viewed in real time. This sensor setup had previously been validated against the Fastrak motion sensor system [22, 23]. Physiotherapists had a 3-h individual training session to familiarize themselves with the standardized palpation of bony landmarks, sensor placement and sensor protocols. Physiotherapists had to practice the protocol at least twice before the study commenced. Each set of movement tests lasted around 20 min. Sensor testing protocols, namely Lumbar and neck movement protocols, are presented in the Supplementary material, section Sensor testing protocols, available at *Rheumatology* online.

**Fig. 1** Flow diagram for assessments

| | Day One | | | Day Two (10-14 days later) |
|---|---|---|---|---|
| | Rater A | | Rater B | Rater A |
| |  |  |  |  |
| Test | 1st | 2nd | 3rd | 4th |

### Sensor data analysis

The peak angle of each sensor movement was recorded by the ViMove software as the mean peak angle from the three repetitions of each movement. Peak angles for Lu and Cx movements were derived from subtracting the maximum angular movement from the sensors above and below the respective regions. The lumbo-pelvic (LP) ratio was calculated by taking the ratio of maximal pelvic flexion to trunk flexion, presenting it as a percentage [25].

### Sample size and statistical considerations

The sample size estimate was based on our primary aim of assessing reliability using intra-class correlation coefficient (ICC) values. In order to define an anticipated ICC of 0.8 with a confidence interval of $\pm$ 0.1, a sample size of 40 was selected [26]. The scale from Bland and Altman was used in the classification of reliability (0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, $\geq$0.81 excellent). Inter-rater, intra-rater and test–retest intra-ICCs were calculated to determine reliability [26–28]. Reliability tests were applied to the values for peak range of movement, and the LP ratio. The two-way random effects, single rater, absolute agreement model for inter-rater, intra-rater and test–retest ICCs were used [27, 29]. SPSS v23 was employed for statistical analysis.

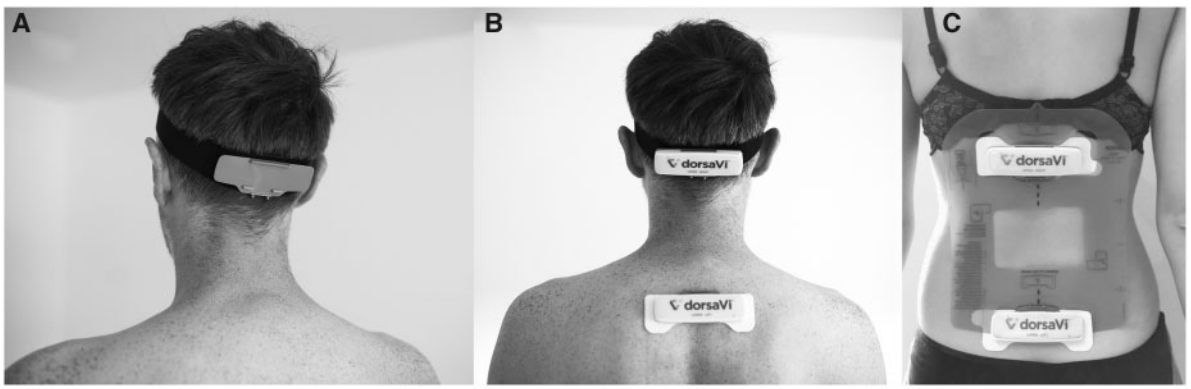
### Data transformation

The ViMove software processes orientation quaternions to calculate angles between IMU sensors. This software also applies filtering and error correction resulting in kinematic data output saved in separate spreadsheet files for each movement test.

The maximum angles at the limits of movement are identified automatically in the software, but we also checked these values manually from spreadsheet data. We did not find this to be a significant source of error.

Normalized scales allow clinicians to quickly assess the severity of mobility restriction without knowing the normal ranges for each movement, and in contrast to range of movement (ROM), the values increase in value

**Fig. 2** Placement of IMU sensors



from 0 to 10 with increasing limitation of movement. Each movement is converted into the same scale even though the range of movement may be quite different. This is widely used in the BASMI, where the raw test results are transformed into normalized scales using the $BASMI_{lin}$ formulae [24]. Each sensor-based movement test result was therefore converted into a normalized index using a similar methodology to that used for the $BASMI_{1lin}$. Values $<1$ or $>9$ U in the normalized 0–10 index were taken as indications of potential floor or ceiling effects, taking into consideration the average change in BASMI scores reported following treatment with biologic drugs [30].

The mean of the normalized scores for each set of movements in each region was reported as the regional composite score for the cervical (Cx), lumbar (Lu) and lumbopelvic (LP) regions. Two further composite scores were developed using the mean of the cervical score (Cx ASMI) and either the lumbar or lumbo-pelvic score (Lu or LPASMI). Each movement was allocated equal 'weight' within the composite IMU-ASMI score. The overall IMU-ASMI score was correlated with the $BASMI_{lin}$ and the BASFI. The intra-rater, inter-rater and test–retest ICCs for these composite scores were calculated. Bland–Altman plots were prepared to identify any systematic difference between the measurements or possible outliers, and to calculate the smallest detectable difference using 95% CIs (mean $\pm$ 1.96 $\times$ S.D. of the mean difference between status scores). The standard error of measurement (SEM) was calculated as follows: SEM $=$ S.D. $\times \sqrt{}$(1 – ICC), with S.D. representing the pooled (two measurements) S.D. of the measure. The smallest detectable change (SDC) is the magnitude of change necessary to provide confidence that a change is not the result of random variation or measurement error, and it is calculated as follows: SDC $=1.96 \times$ SEM $\times \sqrt{}2$ [31].

**Table 1** Descriptive characteristics of study participants ($n = 40$)

| | Mean (S.D.) | Range |
|---|---|---|
| Age, years | 48 (13.4) | 27–71 |
| Disease duration, years | 13 (10.9) | 1–44 |
| BMI, kg/m$^2$ | 27.7 (5.0) | 17.7–39.6 |
| Height, cm | 171.8 (9.6) | 147–190 |
| BAS-G (0–10) | 4.2 (2.8) | 0.3–9.3 |
| BASDAI (0–10) | 4.5 (2.6) | 0–9.9 |
| BASFI (0–10) | 4.6 (3.1) | 0.1–9.7 |
| $BASMI_{lin}$ (0–10) | 5.0 (1.9) | 0.7–8.2 |

$BASMI_{lin}$: Bath Ankylosing Spondylitis Metrology Index (linear).

## Results

### Demographics

The group was comprised of 40 participants, 29 (72.5%) of whom were men. The mean age was 48 (27–71) years, and average disease duration of 13 (1–45) years (Table 1). There was a wide range of disease severity, as reflected in the wide range of BASG, BASDAI, BASFI and $BASMI_{lin}$ values. There was no change in patient or physician-reported disease status or in medication usage in any participant between the first and second study days.

No participant reported side effects from shaving/wearing the sensors, and there were no withdrawals from the study. One participant was not able to complete the second visit due to work commitments, so the 'test–retest' analysis was based on the remaining 39 participants.

The ROM for each measurement using IMU sensors and conventional metrology is shown in Table 2. The range of normalized scores for each movement is shown in Table 3.

### Validity of IMU movement tests

IMU movement tests are reported in angles (Table 2) and can be normalized to provide a global mobility index, providing insights as to which movements are most affected. Overall, 53% of the restriction in the Lu spine was due to limited lateral flexion (23–100%); 27% to limited rotation (0–53%) and 20% due to limited flexion/extension (0–53%). There was considerable variation within individuals regarding the movement with the greatest limitation. With sensor testing, the relative contribution of pelvic and Lu movement to flexion becomes clear: this study showed clinically significant variation in LP patterns. Movements measured by the trunk IMU correlated better with BASFI than 'Lu' movements (Table 2). Two of five BASMI components correlated closely with BASFI ($r > 0.7$) (Table 3). Cx rotation by goniometry correlated strongly with the IMU test ($r = 0.85$). Lu lateral flexion by IMU correlated strongly with the tape measure method ($r = 0.84$). Correlations between Schober's test/Lu IMU-Anterior Flexion/Extension and between Tragus to wall test/Cx IMU-Anterior Flexion/Extension were only moderate ($r = 0.62$ and $r = 0.65$, respectively). Detailed correlation matricies are presented in Supplementary Tables S1 and S2, available at *Rheumatology* online. The CxLP-ASMI and CxLu-ASMI correlated closely with the BASMI ($r = 0.88$ and $r = 0.85$, respectively).

### Reliability of IMU movement tests

Each movement in the protocol was repeated three times without moving the sensors. The ICC for the reliability of the peak ROM estimate was 0.98 overall, 0.99 if the first set of movements was discarded.

The reliability of using combined left/right or flexion/extension movements ('full-arc') or measurements from the

**Table 2** Range of spinal movement in study participants ($n = 40$)

| Method | Movement test | Mean (range) | BASFIcorrelation |
|---|---|---|---|
| Cervical region IMU | Flexion+extension (deg) | 77.5 (5.0–131.0) | −0.5 |
| | Lateral flexion L+R (deg) | 46.1 (3.0–94.0) | −0.4 |
| | Rotation L+R (deg) | 104.0 (11.7–184.3) | −0.6 |
| LP region IMU | Flexion+extension (deg) | 94.9 (36.3–152.0) | **−0.7** |
| | Lateral flexion L+R (deg) | 31.9 (4.3–73.3) | −0.5 |
| | Rotation L+R (deg) | 27.7 (0–65.7) | **−0.7** |
| Lu region IMU | Flexion+extension (deg) | 47.1 (5.3–92.0) | −0.5 |
| | Lateral flexion L+R (deg) | 23.9 (3.0–61.3) | −0.4 |
| | Rotation L+R (deg) | 17.5 (0–42.7) | **−0.7** |
| Conventional metrology | Side flexion L+R (cm) | 19.7 (4.0–41.0) | −0.6 |
| | Tragus to wall distance (cm) | 16.4 (9.8–24.4) | −0.4 |
| | Modified Schöbers (cm) | 3.6 (0.7–7.3 | −0.4 |
| | Intermalleolar distance (cm) | 70.3 (25.5–121.7) | **−0.7** |
| | Cervical rotation L+R (deg) | 87.9 (10.7–170.0) | **−0.7** |
| | Chest expansion (cm) | 3.9 (1.5–9.7) | −0.4 |

LP region: the orientation angle from the upper L1 sensor to the ground, representing both Lu and pelvic movement. Lu region: the angle between the L1 and Sacrum sensors. Strong correlation −0.7 or more shown in bold.

**Table 3** Normalized indices for BASMI and IMU measurements

| Method | Movement test | Mean (range) | Flooreffect (*n*) | Ceiling effect(*n* ) | BASFI correlation |
|---|---|---|---|---|---|
| Cervical IMU | Flexion+extension | 3.0 (0–9.9) | **9** | 1 | 0.5 |
| | Lateral flexion | 4.1 (0–9.4) | **6** | 1 | 0.4 |
| | Rotation | 3.4 (0–9.7) | **8** | 2 | 0.6 |
| LP IMU | Flexion+extension | 4.8 (0.4–9.8) | 2 | 2 | **0.7** |
| | Lateral flexion | 6.0 (0.6–9.8) | 1 | 2 | 0.5 |
| | Rotation | 8.0 (6.0–9.7) | 0 | **9** | 0.6 |
| Lu region IMU | Flexion+extension | 3.7 (0.1–9.2) | 2 | **8** | 0.6 |
| | Lateral flexion | 6.1 (0–9.7) | 1 | **7** | 0.5 |
| | Rotation | 4.9 (0–8.8) | 4 | 0 | **0.7** |
| Conventionalmetrology | Side flexion | 5.3 (0–9.1) | 2 | 1 | 0.6 |
| | Tragus to wall | 2.9 (0.8–4.3) | 1 | 0 | 0.4 |
| | Schöber's test | 3.6 (0.6–9.9) | 2 | 3 | 0.4 |
| | Intramalleolar distance | 5.2 (0.4–9.8) | 4 | 3 | **0.7** |
| | Cervical rotation | 5.2 (0–10) | 1 | 3 | 0.6 |
| IMU regionalASMIs | Cx region ASMI | 3.50 (0–9.7) | 3 | 1 | 0.5 |
| | Lu region ASMI | 4.59 (0.1–9.4) | 0 | 0 | **0.7** |
| | LP-ASMI | 4.40 (1.3–6.5) | 2 | 4 | **0.7** |
| | Cx + Lu ASMI | 4.04 (0.1–9.3) | 0 | 0 | **0.7** |
| | CxLP-ASMI | 3.95 (0.6–7.5) | 4 | 1 | **0.7** |
| $BASMI_{lin}$ | CxLP + hips | 4.83 (1.2–8.4) | 1 | 2 | **0.7** |

Potential ceiling/floor effect >6/40 in bold; correlation coefficient ≥0.7 in bold. IMU: inertial measurement unit.

midline ('half-arc') was compared. The reliability of full-arc movements was slightly higher (Supplementary Table S3, available at *Rheumatology* online), so the combined 'full-arc' movements were used in all subsequent calculations.

The reliability results for IMU and conventional movement tests are shown in Table 4. The intra-rater, inter-rater and test–retest reliability for all the IMU Cx measurements were in the 'good to excellent' range of reliability (ICCs >0.8), but LP and Lu measurements showed slightly lower reliability, particularly the Lu tests. The Lu values are derived by subtracting movement at the pelvic sensor from that at the upper Lu sensor, but it is important to be aware that the pelvic sensor did move significantly in most participants. The conversion of raw angles to normalized indices did not have any effect on reliability (data not shown). No difference was found between intra-rater and inter-rater reliability. Three of the six conventional tests showed good to excellent reliability, but the reliability of chest expansion measurement was particularly poor. Test–retest reliability was generally lower than intra-rater and inter-rater reliability for conventional testing.

TABLE 4 Reliability of IMU and conventional movement tests (ICCs)

| Method | Region/test | Intra-rater ICC[a] | Inter-rater ICC[a] | Test–retest | |
|---|---|---|---|---|---|
| | | | | ICC[a] | SDC95[b] |
| IMU sensor for individual movements | Cervical | | | | |
| | Flexion+extension (deg) | **0.95** | **0.94** | **0.92** | 26.1 |
| | Rotation (deg) | **0.97** | **0.97** | **0.96** | 21.5 |
| | Lateral flexion (deg) | **0.83** | **0.96** | **0.84** | 27.1 |
| | LP | | | | |
| | Flexion+extension (deg) | **0.97** | **0.92** | **0.91** | 23.9 |
| | Rotation (deg) | **0.84** | **0.94** | **0.92** | 18.6 |
| | Lateral flexion (deg) | **0.80** | 0.75 | 0.82 | 11.4 |
| | Lu | | | | |
| | Flexion+extension (deg) | **0.89** | 0.76 | 0.71 | 23.8 |
| | Rotation (deg) | **0.90** | **0.95** | **0.89** | 16.0 |
| | Lateral flexion (deg) | 0.78 | 0.74 | 0.76 | 13.7 |
| Regional Composite IMU scores | Cx-ASMI: units | **0.97** | **0.98** | **0.97** | 1.28 |
| | Lu-ASMI: units | **0.90** | **0.90** | **0.94** | 1.83 |
| | LP-ASMI: units | **0.91** | **0.94** | **0.95** | 1.17 |
| | CxLu-ASMI: units | **0.96** | **0.98** | **0.96** | 1.10 |
| | CxLP-ASMI: units | **0.96** | **0.99** | **0.97** | **0.83** |
| Conventional | Tragus to wall distance (units) | **0.96** | **0.93** | **0.82** | 3.0 |
| | Intermalleolar distance (units) | **0.91** | **0.94** | **0.83** | 2.93 |
| | Cervical rotation (units) | **0.96** | **0.91** | 0.79 | 3.3 |
| | Modified Schöber's test (units) | **1.00** | 0.68 | 0.73 | 3.7 |
| | Lateral flexion (units) | **0.94** | **0.96** | **0.91** | 2.1 |
| | Chest expansion (units) | 0.41 | 0.32 | 0.57 | 4.5 |
| $BASMI_{lin}$ Composite | CxLP: units | **0.97** | **0.98** | **0.96** | **0.91** |

[a]ICC (1, 3) two-way random effects, absolute agreement, single rater. ICCs >0.80 in bold. [b]SDC based on 95% CI < 1 U in bold.

All the regional IMU-ASMI scores showed excellent reliability, particularly the 'Cervico-LP-ASMI' which compares most closely to the BASMI. The reliability of both IMU and conventional movement tests improves when combined into composite indices. Researchers can select the regional mobility score most relevant to their study bearing in mind that the reliability of Lu scores is slightly lower. Bland–Altman graphs were scrutinized for each movement test (graphs not shown). There was no trend towards worse reliability with reduced range of movement. The SDC95 values were comparable or superior to conventional tests, which would suggest that the responsiveness to change of the sensor mobility scores are likely to be superior to conventional tests.

The mean lumbopelvic ratio (LPR) during flexion was 52%, but this varied widely from 7.4% to 98.0%. Six participants had mostly Lu movement (LPR <35%), and eight were pelvic dominant with an LPR >65%. Lu restriction is a characteristic feature in axSpA, but hip arthritis is also relatively common. Five of six participants with severely restricted pelvic movement also had a reduced intermalleolar distance. The intra-rater ICC for LPR ratio measurement was 0.90, inter-rater ICC 0.84, test–retest ICC 0.79.

## Discussion

This study demonstrates that IMU sensor-based measurements in axSpA show strong validity and reliability. This method has the potential to replace conventional measurement tests in clinical practice. We expected reliability in the Lu spine to be greater than in the Cx spine (due to better skin fixation) but the opposite was true. The results in the Cx spine suggest that the 'technical' reliability of sensor measurements was excellent, while in the Lu spine most of the variability was due to 'biological' factors due to the complexities of 'compound' Lu and pelvic movement. The CxLP-ASMI minimizes this variability by ignoring pelvic movement, but the CxLu-ASMI isolates Lu movement and correlates better with Schober's test. Both measures can be reported from a single test.

As expected, Cx rotation measured by sensors was strongly correlated to goniometry, as were the lateral flexion tests by sensor and tape measure methods. The BASMI and CxLP-ASMI were closely correlated ($r = 0.85$).

Of all the patient-reported outcome measures, the BASMI test usually correlates most closely with the BASFI [24]. This was also true of the IMU-ASMI and both measures correlated quite closely with BASFI ($r = 0.7$ for each).

This spinal sensor protocol enables the clinicians to isolate segmental spinal movements within the Cx, Lu and pelvic regions. The LP ratio in our study group covered a surprisingly wide range when compared with previously reported data from normal controls and people with chronic low back pain [32–34]: this aspect of spinal mobility merits further study in axSpA patients. Of the eight participants who were found to have severely restricted pelvic movement, all but two also had a significant reduction in intermalleolar distance, suggesting that the inclusion of pelvic sensor data gives an important insight into the pelvic contribution.

The test–retest reliability of individual Cx movement tests was good to excellent (ICCs $>0.8$), superior to those reported by Theobald *et al*. [21]. Lu movement tests had slightly lower test-reliability (ICCs $>0.7$), similar to the findings reported by Ronchi *et al*. [22] and Laird *et al*. [35] using the same sensor setup.

Combining the right and left or flexion/extension movements improved reliability, probably because it is difficult for assessors to appraise the return to the exact midline point. Measuring the full arc of movement was also shown to be more reliable than half arc movements in a recent study of Cx spine mobility [36]. There was no difference between intra-rater and inter-rater reliability. It was surprising to find that Cx movement tests were more reliable than Lu spine movement tests, as the sensors were not as firmly attached to the skin as in the Lu tests. This suggests that the variability in Lu measurements was due to biological variability rather than sensor error. Laird suggested that it was due to inherent variability in the 'LP rhythm', which was also observed in our study [35, 37, 38]. The test–retest reliability of conventional spinal mobility tests was excellent for side flexion (ICC $>0.9$), good for tragus to wall and intermalleolar distance tests (ICC $> 0.8$) but $<0.8$ for the key tests of Cx rotation and modified Schöber's test. Garrido-Castro *et al*. [17] has previously noted poor reliability for Schöber's test, side flexion and Cx rotation. In that study, it was shown that movement tests using the UCOTrack motion capture method showed uniformly excellent levels of inter-rater reliability apart from frontal spinal flexion.

Converting raw movement angles into normalized scales does not negatively impact test–retest reliability. This stage is an important intermediate step in developing a composite spinal mobility score that further improves reliability and reduces the potential for floor/ceiling artefact. It allows restrictions in different planar movements to be compared without further adjustments. For instance, in this study 53% of the composite Lu index was due to limited lateral flexion (range 23–100%); 27% to limited rotation (0–53%) and 20% due to limited flexion/extension (0–53%). There was considerable variation within individuals as to which movement showed the greatest limitation.

The reliability of the regional composite indices (Cx, Lu, LP, Cervico-Lu and Cervico-LP) was clearly superior to that of the individual components and showed fewer floor/ceiling effects (Table 4). The regional indices provide insights as to which regions are most affected. For instance, in this group of individuals, 68% (range 42–100%) of the CxLP-ASMI was due to LP limitation, while 32% (range 0–57%) was due to Cx limitation. The reliability of the IMU sensor-based ASMI reported here is similar to that reported for the motion-capture based UCOASMI [17].The limitations of this study include a probable underestimation of the trunk rotation angle. We used a trans-Lu sensor positioning as at that time there was no validated protocol for measurement across the whole thoracic spine. Moreover, the precision of normalized scores would be improved by referencing the range of movements of a larger, age-adjusted, normal population [39, 40].

## Conclusion

This study has demonstrated that an IMU sensor-based method of measuring spinal mobility in axSpA is valid, reliable and able to give a detailed and reliable 'snapshot' of spinal mobility in different dimensions and over different regions of the spine. These tests correlate both with conventional mobility tests and with physical function. Physiotherapists or other trained health professionals can perform the test in a standard clinic setting equipped with sensors and a laptop. The clinician is presented with a range of maximum angles of movement in the Cx and Lu spine from which normalized indices of spinal mobility can be derived.

## Acknowledgements

The authors acknowledge the willing collaboration of all the patients who were partners in this research. They also acknowledge the hard work and expertise of Eithne Boyle, Ann-Marie Conlon and Stephanie Keys. P.M.M. is supported by the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the (UK) National Health Service (NHS), the NIHR or the (UK) Department of Health.

Preliminary data included in this publication was published at the ACR Annual Meeting 2018.

*Funding*: This work was supported by a project grant from the Foundation for Research in Rheumatology (FOREUM). The funders had no role in study design, data collection, analysis, decision to publish or preparation of the manuscript. Sensors were purchased from DorsaVi ltd. under a research agreement. DorsaVi had no input in the analysis or interpretation of data.

*Disclosure statement*: P.V.G. reports honoraria from Celgene; consulting fees from Pfizer and Genomics Medicine Ireland; involvement in a clinical trial funded by Abbvie. P.M.M. reports honoraria/consulting fees from Abbvie, BMS, Celgene, Janssen, MSD, Novartis, Pfizer, Roche and UCB. The other authors have declared no conflicts of interest.

## Supplementary data

Supplementary data are available at *Rheumatology* online.

## References


1 van der Heijde D, Calin A, Dougados M *et al*. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. Assessments in Ankylosing Spondylitis. J Rheumatol 1999;26:951–4.

2 EMEA. *Committee for Medicinal Products for Human Use (Chmp) Guidance on Clinical Investigation of Medicin Guideline on Clinical Investigation of Medicinal Products for the Treatment of Ankylosing Spondylitis*. 2005:1–9. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-treatment-axial-spondyloarthritis-revision-1_en.pdf

3 Jenkinson TR, Mallorie PA, Whitelock HC *et al*. Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index. J Rheumatol 1994;21:1694–8.

4 Madsen OR, Hansen LB, Rytter A, Suetta C, Egsmose C. The Bath metrology index as assessed by a trained and an untrained rater in patients with spondylarthropathy: a study of intra- and inter-rater agreements. Clin Rheumatol 2009;28:35–40.

5 Jauregui E, Conner-Spady B, Russell AS, Maksymowych WP. Clinimetric evaluation of the bath ankylosing spondylitis metrology index in a controlled trial of pamidronate therapy. J Rheumatol 2004;31:2422–8.

6 Machado P, Landewe R, Braun J *et al*. A stratified model for health outcomes in ankylosing spondylitis. Ann Rheum Dis 2011;70:1758–64.

7 Sieper J, Rudwaleit M, Baraliakos X *et al*. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. Ann Rheum Dis 2009;68 (Suppl 2):ii1–44.

8 Maksymowych WP, Mallon C, Richardson R *et al*. Development and validation of the Edmonton Ankylosing Spondylitis Metrology Index. Arthritis Rheum 2006;55:575–82.

9 Edmondston SJ, Aggerholm M, Elfving S *et al*. Influence of posture on the range of axial rotation and coupled lateral flexion of the thoracic spine. J Manip Physiol Ther 2007;30:193–9.

10 Cuesta-Vargas AI, Galán-Mercant A, Williams JM. The use of inertial sensors system for human motion analysis. Phys Ther Rev 2010;15:462–73.

11 Tsang SMH, Szeto GPY, Lee R. Movement coordination and differential kinematics of the cervical and thoracic spines in people with chronic neck pain. Clin Biomech 2013;28:610–7.

12 Burnett AF, Barrett CJ, Marshall RN, Elliott BC, Day RE. Three-dimensional measurement of lumbar spine kinematics for fast bowlers in cricket. Clin Biomech 1998;13:574–83.

13 Jordan K, Haywood KL, Dziedzic K *et al*. Assessment of the 3-dimensional fastrak measurement system in measuring range of motion in ankylosing spondylitis. J Rheumatol 2004;31:2207–15.

14 Milne AD, Chess DG, Johnson JA, King GJW. Accuracy of an electromagnetic tracking device: a study of the optimal operating range and metal interference. J Biomech 1996;29:791–3.

15 Garrido-Castro JL, Medina-Carnicer R, Schiottis R *et al*. Assessment of spinal mobility in ankylosing spondylitis using a video-based motion capture system. Man Ther 2012;17:422–6.

16 Garrido-Castro JL, Escudero A, Medina-Carnicer R *et al*. Validation of a new objective index to measure spinal mobility: the University of Cordoba Ankylosing Spondylitis Metrology Index (UCOASMI). Rheumatol Int 2014;34:401–6.

17 Garrido-Castro JL, Curbelo R, Mazzucchelli R *et al*. High reproducibility of an automated measurement of mobility for patients with axial spondyloarthritis. J Rheumatol 2018;45:1383–8.

18 Laird R. A, Kent P, Keating JL. Modifying patterns of movement in people with low back pain – does it help? A systematic review. BMC Musculoskelet Disord 2012;13:169.

19 Kent P, Laird R, Haines T. The effect of changing movement and posture using motion-sensor biofeedback, versus guidelines-based care, on the clinical outcomes of people with sub-acute or chronic low back pain-a multi-centre, cluster-randomised, placebo-controlled, pilot trial. BMC Musculoskelet Disord 2015;16:131.

20 Mjøsund HL, Boyle E, Kjaer P *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 Musculoskelet Disord 2017;18:1–9.

21 Theobald PS, Jones MD, Williams JM. Do inertial sensors represent a viable method to reliably measure cervical spine range of motion? Man Ther 2012;17:92–6.

22 Ronchi AJ, Lech M, Taylor NF, Cosic I. A reliability study of the new Back Strain Monitor based on clinical trials. Conf Proc IEEE Eng Med Biol Soc 2008;2008:693–6.

23 Charry E, Umer M, Taylor S. Design and Validation of an Ambulatory Inertial System for 3-D Measurements of Low Back Movements. In: *Proceedings of the 2011 7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, *ISSNIP 2011*; 2011:58–63. doi: 10.1109/ISSNIP.2011.6146618.

24 van der Heijde D, Landewé R, Feldtkeller E. Proposal of a linear definition of the Bath Ankylosing Spondylitis Metrology Index (BASMI) and comparison with the 2-step and 10-step definitions. Ann Rheum Dis 2008;67:489–93.

25 Kent P. Comparing lumbo-pelvic movement in people with and without back pain: a systematic review. BMC Musculoskelet Disord 2014;15:13.

26 De Vet HCW. Measurement in Medicine. Cambridge: Blackwell's, 2011.

27 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.

28 Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–63.

29 Auleley G-R, Benbouazza K, Spoorenberg A *et al*. Evaluation of the smallest detectable difference in outcome or process variables in ankylosing spondylitis. Arthritis Rheum 2002;47:582–7.

30 Sepriano A, Regel A, Van Der Heijde D *et al*. Efficacy and safety of biological and targeted-synthetic DMARDs: a systematic literature review informing the 2016 update of the ASAS/EULAR recommendations for the management of axial spondyloarthritis. RMD Open 2017;3:e000396.

31 Šerbetar I. Establishing some measures of absolute and relative reliability of a motor tests. Croat J Educ 2015;17: 37–48.

32 Burton AK, Tillotson KM. Reference values for "normal" regional lumbar sagittal mobility. Clin Biomech 1988;3: 106–13.

33 Laird RA, Keating JL, Kent P. Subgroups of lumbo-pelvic flexion kinematics are present in people with and without persistent low back pain. BMC Musculoskelet Disord 2018;19:13.

34 Vazirian M, Van Dillen L, Bazrgari B. Lumbopelvic rhythm during trunk motion in the sagittal plane: a review of the kinematic measurement methods and characterization approaches. Phys Ther Rehabil 2016;3:5.

35 Laird RA, Kent P, Keating JL *et al*. How consistent are lordosis, range of movement and lumbo-pelvic rhythm in people with and without back pain? BMC Musculoskelet Disord 2016;17:403.

36 Raya R, Garcia-Carmona R, Sanchez C *et al*. An inexpensive and easy to use cervical range of motion measurement solution using inertial sensors. Sensors 2018;18:2582–17.

37 Macmanus C, Dankaerts W, Sullivan KO. Using wireless technology to monitor lumbo-pelvic kinematics in rowing. Physiotherapy Practice and Research 2013; 34:113–21.

38 O'Sullivan K, O'Sullivan L, Campbell A, O'Sullivan P, Dankaerts W. Towards monitoring lumbo-pelvic posture in real-life situations: concurrent validity of a novel posture monitor and a traditional laboratory-based motion analysis system. Man Ther 2012;17:77–83.

39 Marques ML, Ramiro S, Goupille P *et al*. Measuring spinal mobility in early axial spondyloarthritis: does it matter? Rheumatology 2019;58:1597–606.

40 Ramiro S, Van Tubergen A, Stolwijk C *et al*. Reference intervals of spinal mobility measures in normal individuals: the mobility study. Ann Rheum Dis 2015;74: 1218–24.