Anne F. Mannion
Katrin Knecht
Gordana Balaban
Jiri Dvorak
Dieter Grob

A new skin-surface device for measuring the curvature and global and segmental ranges of motion of the spine: reliability of measurements and comparison with data reviewed from the literature

A. F. Mannion (✉) · K. Knecht
G. Balaban · J. Dvorak · D. Grob
Spine Unit, Schulthess Klinik,
Lengghalde 2, 8008 Zürich, Switzerland
Tel.: +41-1-3857584,
Fax: +41-1-3857590,
e-mail: afm@kws.ch

A. F. Mannion
Institute of Physical Medicine
and Rehabilitation, University Hospital,
Zürich, Switzerland

Abstract There is an increasing awareness of the risks and dangers of exposure to radiation associated with repeated radiographic assessment of spinal curvature and spinal movements. As such, attempts are continuously being made to develop skin-surface devices for use in examining the progression and response to treatment of various spinal disorders. However, the reliability and validity of measurements recorded with such devices must be established before they can be recommended for use in the research or clinical environment. The aim of this study was to examine the reliability of measurements using a newly developed skin-surface device, the Spinal Mouse. Twenty healthy volunteers (mean age 41±12 years, nine males, 11 females) took part. On 2 separate days, spinal curvature was measured with the Spinal Mouse during standing, full flexion, and full extension (each three times by each of two examiners). Paired *t*-tests, intraclass correlation coefficients (ICC), and standard errors of measurement (SEM) with 95% confidence intervals were used to characterise between-day and interexaminer reliability for: standing sacral angle, lumbar lordosis, thoracic kyphosis, and ranges of motion (flexion, extension) of the thoracic spine, lumbar spine, hips, and trunk. The between-day reliability for segmental ranges of flexion was also determined for each motion segment from T1-2 to L5-S1. The majority of parameters measured for the 'global regions' (thoracic, lumbar, or hips) showed good between-day reliability. Depending on the parameter of interest, between-day ICCs ranged from 0.67 to 0.92 for examiner 1 (average 0.82) and 0.57 to 0.95 for examiner 2 (average 0.83); for 70% of the parameters measured, the ICCs were greater than 0.8 and generally highest for the lumbar spine and whole trunk measures. For lumbar spine range of flexion, the SEM was approximately 3°. The ICCs were also good for the interexaminer comparisons, ranging from 0.62 to 0.93 on day 1 (average 0.81) and 0.70 to 0.94 on day 2 (average 0.86), although small systematic differences were sometimes observed in their mean values. The latter were still evident even if both examiners used the same skin markings. For segmental ranges of flexion, the ICCs varied between vertebral levels but overall were lower than for the global measures (average for all levels in all analyses, ICC 0.6). For each examiner, the average between-day SEM over all vertebral levels was approximately 2°. For 'global' regions of the spine, the Spinal Mouse delivered consistently reliable values for standing curvatures and ranges of motion which compared well with those reported in the literature. This suggests that the device can be reliably implemented for in vivo studies of the

sagittal profile and range of motion of the spine. As might be expected for the smaller angles being measured, the segmental ranges of flexion showed lower reliability. Their usefulness with regard to the interpretation of individual results and the detection of 'real change' on an individual basis thus remains questionable. Nonetheless, the group mean values showed few between-day differences, suggesting that the device may still be of use in providing clinically interesting data on segmental motion when examining groups of individuals with a given spinal pathology or undergoing some type of intervention.

Introduction

With increasing awareness of the dangers of exposure to radiation associated with the repeated radiographic assessment of spinal curvature and spinal movements [31], attempts are continuously being made to develop skin-surface devices for use in examining the progression and response to treatment of various spinal disorders. A number of devices employing different methods/techniques of measurement are currently available for the noninvasive assessment of spinal movements – ranging from the simple tape measure to computerized motion analysis systems. The latter (e.g. Fastrak, Isotrak, Zebris CMS, and CA6000) most commonly indicate the standing curvature and range of motion of a given section of the whole spine (e.g. the lumbar region) using various movement sensors affixed to the skin surface at positions believed to correspond to the underlying vertebrae, as determined by prior palpation and skin marking [8, 27, 35, 37, 41, 46, 48]. These computerised motion analysis devices offer the additional advantage of being able to monitor and record continuously the changing curvature of the spine, thereby allowing both the pattern and extent of movement to be assessed not only during range-of-motion testing but also during the performance of given activities (e.g. bending and lifting movements, locomotion) [1, 8, 9, 51].

In general, most of these motion analysis devices have been shown to be reliable [14, 27]. However, they also have certain drawbacks. Palpation of the precise landmarks for placement of the sensors as well as preparation of the skin for their firm attachment can both be time-consuming. Inaccuracies in the angles measured can arise not only as a result of skin movement unrelated to underlying vertebral movement (a problem common to all skin-surface devices) but also if true contact, i.e. complete apposition, between the sensors and the skin is not maintained throughout the testing procedure. Thus, the firm and stable attachment of the measuring device to the individual's back plays a crucial role in obtaining reliable measures. In the case of devices that are attached to the patient by means of straps or belts, slippage may be a problem if considerable movement occurs between standing and end ranges of motion; with sensors that are attached to the skin with adhesive/sticky tape, problems may arise if the contact between the sensors and the skin deteriorates, for example if the patient begins to sweat.

Devices that are not attached to the person per se but rather are placed upon the back for measurement once the given posture has been adopted include kyphometers [40], goniometers [40], inclinometers [53], and flexicurves [5, 43, 45]. Kyphometers, goniometers, and dual inclinometers have the disadvantage that only one global region of the spine can be monitored at a time (e.g. thoracic or lumbar spine or sacral tilt), whilst the data analysis from flexicurve measurements is rather cumbersome for use in routine practice, when rapid feedback of the results is required.

Two computer-aided skin-surface devices have recently been developed to measure the spinal curvature of each of the main global regions of the spine (lumbar, thoracic, sacral) as well as that of the motion segments from T1-2 to L5-S1. One of these is based on a pen-pointer device fitted with ultrasonic transmitters positioned sequentially over each of the spinal processes to detect their 3D position (Zebris, Germany). The second is a wheeled device housing accelerometers which records distance and changes of inclination with regard to the plumb line as it is rolled along the length of the spine (effectively, an electronic inclinometer) (Spinal Mouse) (Idiag, Voletswil, Switzerland). Although both devices have been on the market for some years, to the authors' knowledge there are currently no independent reports in the peer-reviewed literature concerning their reliability or accuracy.

The aim of the present study was to assess the reliability of one of these types of device, the Spinal Mouse, with regard to measures of standing sagittal curvature, global range of motion of the hips and lumbar and thoracic regions of the spine, and the range of flexion (ROF) of individual motion segments of the whole spine from T1-2 down to L5-S1.

Materials and methods

Subjects

Twenty healthy volunteers agreed to participate in the study. There were nine males (45.4±7.7 years, 1.81±0.09 m tall, 87.6±16.8 kg, and body mass index, or BMI, 26.5±5.2 kg/m^2) and 11 females (38.2±7.6 years, 1.67±0.10 m, 63.8±7.6 kg, and 22.9±5.7 kg/m^2). None had any low back pain at the time of testing or had done so within the preceding 2 weeks. They were all workers from various medical, technical, and administrative branches of an orthopaedic hospital. Sixty per cent were habitually physically active, whilst 40% did no regular sports. Fifteen per cent declared having very

Fig. 1 Spinal Mouse measurements in standing position, flexion, and extension and the typical output derived from one set of tests
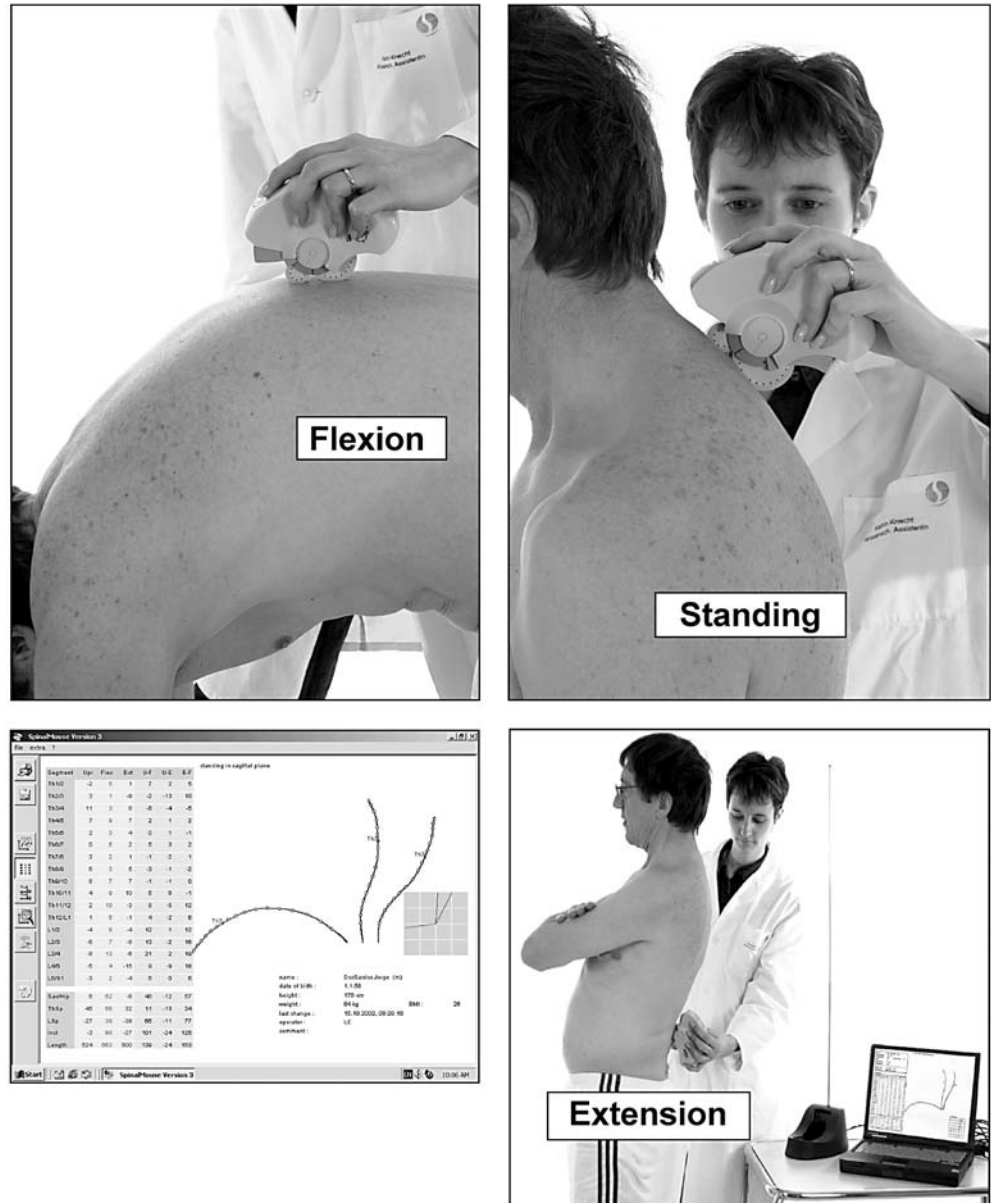
physically demanding jobs, 60% moderately physically demanding jobs, and 25% physically undemanding jobs. Fifty per cent had experienced some back pain in their lives, with the most recent episodes occurring 8–300 months previously; of those, 50% had visited doctors and 30% physiotherapists for treatment. The volunteers gave signed, informed consent to participate. The project was part of a larger study investigating the influence of lumbar surgery on spinal movements which was approved by the local ethics committee.

Equipment

Measures of spinal mobility were made using the Spinal Mouse system, a hand-held, computer-assisted electromechanical device that can be used to measure spinal curvature in various postures [42] (Fig. 1). The device is guided along the midline of the spine (or slightly paravertebrally in particularly thin individuals with prominent spinous processes) starting at the spinous process of C7 and finishing at the top of the anal crease (approximately S3); these landmarks are firstly determined by palpation and marked on the skin surface with a cosmetic pencil. Two rolling wheels follow the contour of the spine, and distance and angle measures are communicated from the device to a base station positioned approximately 1–2 m away and interfaced to a personal computer. Data is sampled every 1.3 mm as the mouse is rolled along the spine, giving a sampling frequency of approximately 150 Hz. (The average total length of the spine is 550 mm and the time required to measure the whole length is 2–4 s; thus, approximately 423 measurements are made over about 3 s). This information is then used to calculate the relative positions of the sacrum and vertebral bodies of the underlying bony spinal column using an intelligent, recursive algorithm.
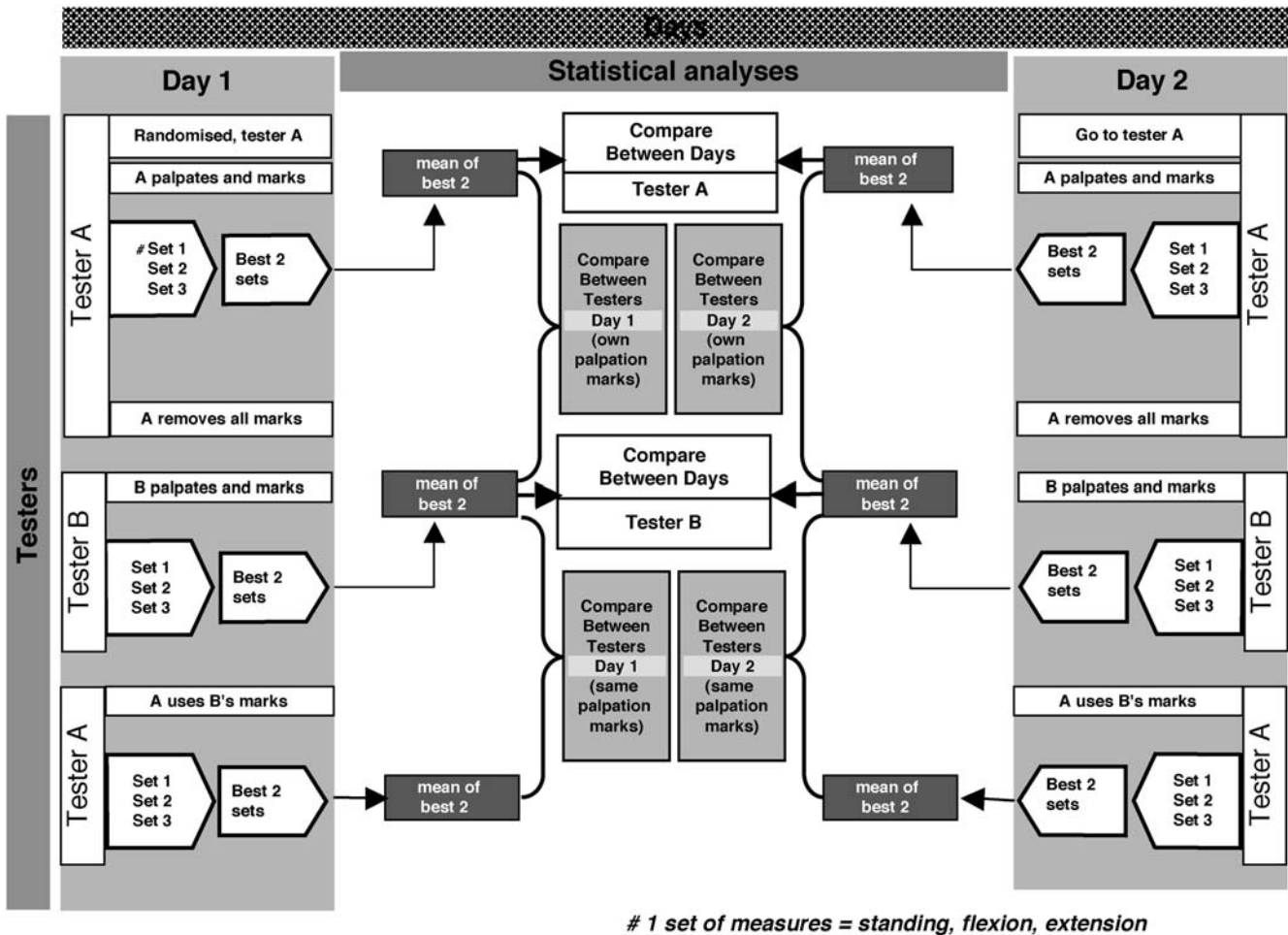
Fig. 2 Protocol for the testing procedures and data analyses. See text for further details

Measurement protocol

The volunteers were randomized to go firstly to examiner 1 or examiner 2 (Fig. 2). The corresponding examiner then palpated the volunteer, marked the landmarks on the skin, and made a set of measurements in the postures described below (one 'set' of measurements will always refer to the three positions of standing, flexion, and extension). Two further sets of measures were then carried out approximately 1–2 min apart. The skin marks were then completely removed, and the volunteer went to the second examiner to be palpated, marked anew, and carry out the same three sets of measures. With the skin marks left on, the patient then returned to the first examiner and performed three further sets of movements, with examiner 1 now using the skin marks placed by examiner 2.

On the 2nd day of testing, at approximately the same time of day for each subject, all the tests described for the 1st day were repeated. The test order was retained to maintain as constant any errors arising from slight increases in mobility with repeated testing (these were expected to be peculiar to the experimental situation of multiple testing and would not therefore contribute to the normal error of measurement). Before starting the study, the examiners discussed and agreed upon the method of palpation and the instructions to be given to the volunteers but did not confer with each other during the measurements themselves.

Test positions

The three test positions adopted for each set of measures comprised:

1. Standing upright (in a relaxed position, focusing on a marker at eye level, feet shoulder width apart, knees straight, arms hanging by the side)
2. Maximal flexion (legs straight, trunk flexed as far as comfortably possible in an attempt to curl the head into the knees, hands gripping the back of the lower leg for stability, if necessary)
3. Maximal extension (legs straight, arms crossed over the front of the body, head in a neutral position, trunk extended as far as comfortably possible)

The positions were first described and demonstrated by the investigator and practiced once by the volunteer before the three sets of measurements in each posture were made. The patient was instructed to move at a speed of his/her choosing and to hold the end position for a few seconds while the measurement was made.

The relevant parameters recorded by the Spinal Mouse in each position were: all the individual motion segment angles (from T1-2 through to L5-S1), thoracic curvature (T1-2 to T11-12), lumbar curvature (T12-L1 to the sacrum), 'hip' (sacral) angle, and trunk angle of inclination (angle subtended between the vertical and a line joining C7 to the sacrum). Determination of these parameters in standing, full flexion, and full extension then allowed calculation of the ranges of flexion and extension for the hips, lumbar spine, thoracic spine, and whole trunk. In addition, the ROF was

Table 1 Between-day reliability for each examiner and interexaminer reliability on each testing day. *CI* confidence interval, *ROF* range of

Position	Between-day reliability							
	Examiner 1					Examiner 2		
	Day 1, mean±SD	Day 2, mean±SD	*P* (day 1 vs day 2)	ICC (day 1 vs day 2) (95% CI)	SEM (day 1 vs day 2) (95% CI)	Day 1, mean±SD	Day 2, mean±SD	*P* (day 1 vs day 2)
Hips								
Standing	21.1±6.2°	21.3±6.9°	0.76	0.88 (0.71–0.95)	2.3 (1.8–3.4)	16.1±6.9°	16.7±7.7°	0.35
ROF	48.9±10.2°	46.9±10.5°	0.16	0.82 (0.59–0.93)	4.4 (3.4–6.5)	49.3±11.3°	47.9±10.0°	0.25
ROE	18.5±9.1°	19.8±10.3°	0.39	0.78 (0.52–0.91)	4.5 (3.5–6.6)	18.7±9.4°	20.6±12.0°	0.40
ROFE	67.3±12.1°	66.7±12.0°	0.75	0.77 (0.50–0.90)	5.8 (4.4–8.6)	67.9±11.5°	68.3±13.5°	0.83
Thoracic spine								
Standing	45.8±6.8°	44.7±9.2°	0.40	0.73 (0.43–0.89)	4.2 (3.2–6.2)	46.7±8.1°	45.8±6.5°	0.31
ROF	24.5±9.1°	25.3±9.9°	0.63	0.67 (0.32–0.86)	5.5 (4.2–8.0)	26.9±11.5°	27.7±9.0°	0.49
ROE	1.1±13.0°	2.1±12.3°	0.63	0.76 (0.47–0.90)	6.2 (4.7–9.1)	0.6±10.3°	2.1±9.4°	0.32
ROFE	25.6±13.6°	27.4±13.3°	0.31	0.83 (0.61–0.93)	5.6 (4.3–8.2)	27.5±12.9°	29.8±12.5°	0.26
Lumbar spine								
Standing	−31.7±7.3°	−32.2±8.4°	0.54	0.90 (0.75–0.96)	2.5 (1.9–3.7)	−32.4±8.2°	−32.7±8.2°	0.74
ROF	63.2±8.0°	64.4±8.8°	0.24	0.85 (0.66–0.94)5	3.2 (2.5–4.7)	67.1±11.0°	67.1±9.4°	0.94
ROE	14.8±9.3°	13.5±10.7°	0.36	0.80 (0.55–0.92)	4.5 (3.5–6.6)	15.4±8.7°	12.9±10.9°	0.10
ROFE	78.1±12.6°	78.0±13.4°	0.93	0.87 (0.69–0.95)	4.7 (3.6–6.8)	82.5±13.9°	80.1±13.9°	0.15
Whole trunk								
Standing	5.2±2.6°	4.8±2.5°	0.29	0.84 (0.64–0.94)	1.0 (0.8–1.5)	0.9±2.6°	0.8±3.1°	0.64
ROF	109.2±10.2°	108.4±11.7°	0.50	0.89 (0.75–0.96)	3.6 (2.7–5.2)	112.0±11.6°	111.9±10.9°	0.93
ROE	32.9±8.9°	33.9±9.2°	0.32	0.88 (0.73–0.95)	3.1 (2.3–4.5)	33.4±8.2°	33.4±8.9°	0.98
ROFE	142.1±14.8°	142.3±16.3°	0.87	0.92 (0.81–0.97)	4.3 (3.3–6.3)	145.4±15.5°	145.3±15.9°	0.94
Mean of all parameters				0.82	4.1			

*P<0.05

determined for each vertebral motion segment (from T1-2 to L5-S1).

Data analysis

Of the three repeated sets of movements performed at any given time (e.g. with one particular examiner or on a given day), only the results of the two 'best' sets were used for further analysis – those in which the greatest whole trunk inclination was achieved during the flexion test (because flexion is the movement most commonly-investigated and of generally greatest interest, and overall trunk flexion was considered to give the best indication of 'maximal effort'.) In the majority of cases, the best two trials were the latter two in the set of three. The following analyses were performed for each of the parameters measured by the Spinal Mouse (Fig. 2):

1. Intraexaminer, between-day reliability (examiner vs himself on 2 days). Comparison of the means of the two best values on day 1 and the two best values on day 2 (analyses done separately for each examiner)
2. Interexaminer reliability (examiner 1 vs examiner 2 on each day). Comparison of the mean of the two best values of examiner 1 on a given day with the mean of the two best values of examiner 2 on the same day, in which each examiner independently palpated and marked the skin of the volunteer
3. Comparison of the means of the two best values of examiner 1 and examiner 2 on the same day, in which the same skin markings were used by both examiners

Statistics

Paired *t*-tests were used to examine group mean differences in the repeated measures for each of the parameters derived from use of the Spinal Mouse. Significance was accepted at the 5% level. As recommended by Perneger [36], no adjustments were made for multiple testing.

The intraclass correlation coefficient (ICC) and standard error of measurement (SEM) (or 'typical error of measurement'), each with 95% confidence intervals, were used to characterise reliability [21]. It has been suggested that, to be useful, a measurement should have an ICC of >0.6 [6]. However, in the present study, the rather more stringent criteria of Currier [7] were adopted: 0.90–0.99=high reliability, 0.80–0.89=good reliability, 0.70–0.79=fair reliability, <0.69=poor reliability.

Results

Global ranges of motion for thoracic spine, lumbar spine, and hips

Intraexaminer between-day reliability

Repetition of the measurements by the same examiner on two different days resulted in no significant difference in

flexin, *ROE* range of extension, *ROFE* range of flexion and extension, *stand* angle in standing

ICC (day 1 vs day 2) (95% CI)	SEM (day 1 vs day 2) (95% CI)	Interexaminer reliability					
		Day 1			Day 2		
		P (examiner1 vs examiner 2)	ICC (examiner 1 vs examiner 2) (95% CI)	SEM (examiner 1 vs examiner 2) (95% CI)	*P* (examiner 1 vs examiner 2)	ICC (examiner 1 vs examiner 2) (95% CI)	SEM (examiner 1 vs examiner 2) (95% CI)
0.93 (0.83–.97)	1.9 (1.5–2.8)	0.00*	0.88 (0.71–0.95)	2.3 (1.8–3.4)	0.00*	0.92 (0.80–.97)	2.1 (1.6–3.1)
0.87 (0.70–0.95)	3.8 (2.9–5.6)	0.55	0.88 (0.71–0.95)	3.8 (2.9–5.6)	0.44	0.85 (0.64-0.94)	4.0 (3.1–5.9)
0.57 (0.17–0.81)	7.0 (5.4–10.3)	0.69	0.67 (0.32–0.86)	5.3 (4.1–7.8)	0.54	0.89 (0.70–0.95)	4.0 (3.0–5.8)
0.70 (0.38–0.87)	6.7 (5.2–10.0)	0.92	0.65 (0.29–0.85)	7.0 (5.3–10.2)	0.23	0.89 (0.73–0.96)	4.3 (3.3–6.2)
0.88 (0.67–0.94)	2.8 (2.1–4.0)	0.33	0.87 (0.70–0.95)	2.7 (2.0–3.9)	0.30	0.83 (0.62–0.93)	3.3 (2.5–4.8)
0.86 (0.67–0.94)	3.9 (2.9–5.6)	0.09	0.83 (0.61–0.93)	4.3 (3.3–6.2)	0.16	0.70 (0.38–0.87	5.2 (3.9–7.5)
0.78 (0.52–0.91)	4.6 (3.5–6.7)	0.80	0.64 (0.27–0.84)	7.0 (5.4–10.3)	0.99	0.79 (0.54–0.92)	5.0 (3.8–7.4)
0.77 (0.51–0.91)	6.0 (4.6–8.8)	0.46	0.62 (0.24–0.83)	8.2 (6.3–12.0)	0.16	0.85 (0.65–0.94)	5.1 (3.9–7.4)
0.92 (0.80–0.97)	2.4 (1.8–3.5)	0.41	0.87 (0.69–0.95)	2.8 (2.1–4.1)	0.52	0.93 (0.83–0.97)	2.2 (1.7–3.2)
0.91 (0.79–0.97)	3.0 (2.3–4.4)	0.00*	0.85 (0.65–0.94)	3.7 (2.8–5.5)	0.02*	0.87 (0.70–0.95)	3.3 (2.5–4.8)
0.78 (0.52–0.91)	4.6 (3.5–6.8)	0.38	0.88 (0.72–0.95)	3.1 (2.3–4.5)	0.58	0.90 (0.75–0.96)	3.5 (2.7–5.1)
0.86 (0.68–0.95)	5.1 (3.9–7.5)	0.01*	0.87 (0.69–0.95)	4.8 (3.7–7.0)	0.20	0.86 (0.68–0.95)	5.0 (3.8–7.3)
0.83 (0.62–0.93)	1.2 (0.9–1.7)	0.00*	0.77 (0.50–0.91)	1.2 (0.9–1.8)	0.00*	0.71 (0.39–0.88)	1.5 (1.2–2.2)
0.95 (0.86–0.98)	2.6 (2.0–3.8)	0.01*	0.93 (0.83–0.97)	2.8 (2.1–4.1)	0.01*	0.87 (0.69–0.95)	4.1 (3.2–6.1)
0.84 (0.63–0.93)	3.4 (2.6–5.0)	0.77	0.87 (0.69–0.95)	3.1 (2.3–4.5)	0.49	0.94 (0.85–0.98)	2.3 (1.7–3.3)
0.94 (0.84–0.97)	4.0 (3.0–5.8)	0.04*	0.90 (0.77–0.96)	4.7 (3.6–6.8)	0.04*	0.93 (0.82–0.97)	4.4 (3.4–6.4)
0.84	4.0		0.81	4.2		0.86	3.7

group mean values for any of the parameters investigated (Table 1). The ICCs ranged between 0.67 and 0.92 for examiner 1 and 0.57 and 0.95 for examiner 2, depending on the parameter of interest. The majority (68%) of ICCs were greater than 0.8 ('good reliability'). The mean between-days ICCs for all parameters were 0.82 for examiner 1 and 0.84 for examiner 2.

For examiner 1, the between-day SEM ranged from 1.0° (95% CI 0.8–1.5) for standing trunk angle to 6.2° (95% CI 4.7–9.1) for the range of extension (ROFE) of the thoracic spine (Table 1). For examiner 2, the between-day SEM ranged from 1.2° (95% CI 0.9–1.7) for standing trunk angle to 7.0° (95% CI 5.4–10.3) for the range of extension (ROE) of the hips (Table 2).

Comparison of the measurements made by the two examiners resulted in a number of significant differences in group mean values (Table 1). Specifically, interexaminer differences were consistently observed (i.e. on each of the days) for standing hip angle, ROF of the lumbar spine and the standing trunk inclination, trunk ROF, and trunk ROFE. However, the corresponding ICCs for each of these parameters were generally high, indicating that the differences represented primarily systematic errors, i.e. one ex-

aminer consistently recorded slightly higher or lower values than the other. Further, the differences were of an order of just a few degrees and very close to the SEM for the corresponding intraexaminer analyses. The ICCs between the measures of the two examiners ranged between 0.62 and 0.93 on day 1 and between 0.70 and 0.94 on day 2, depending on the parameter of interest. The ICCs were over 0.8 (good reliability) for 69% of the parameters on day 1 and 81% of the parameters on day 2. The mean ICCs for all parameters were 0.81 (day 1) and 0.86 (day 2). The interexaminer SEM ranged from 1.2° (95% CI 0.9–1.8) for the standing trunk angle to 7.0° (95% CI 5.3–10.2) for ROE of the hips and ROFE of the thoracic spine (Table 1).

Interexaminer reliability – same skin-markings used by both examiners, day 1

It has been shown that a large part of the interexaminer variability in measurements of spinal mobility arises from differences in identification of the landmarks required for measurement [28]. Thus, the present study also examined

Table 2 Between-day reliability for each examiner and interexaminer reliability on each day for segmental ranges of flexion. *ROF* range of

Vertebral level	Between-day reliability							
	Examiner 1					Examiner 2		
	Day 1, mean±SD	Day 2, mean±SD	*P* (day 1 vs day 2)	ICC, day 1 vs day 2 (95% CI)	SEM (95% CI)	Day 1, mean±SD	Day 2, mean±SD	*P* (day 1 vs day 2)
T1-2	−0.4±6.3°	−0.1±5.1°	0.76	0.76 (0.48–0.90)	2.8 (0.1–4.1)	3.7±5.5°	−0.2±5.4°	0.00*
T2/3	2.6±4.4°	2.4±4.9°	0.84	0.56 (0.16–0.80)	3.1 (2.3–4.5)	3.6±3.2°	4.9±4.5°	0.15
T3-4	−0.1±4.3°	−0.5±3.3°	0.62	0.45 (0.01–0.74)	2.8 (2.2–4.1)	0.3±4.0°	1.0±4.7°	0.34
T4-5	0.5±4.1°	1.2±4.1°	0.33	0.71 (0.39–0.88)	2.2 (1.7–3.3)	1.6±3.4°	1.3±3.7°	0.66
T5-6	2.4±2.9°	2.3±3.7°	0.90	0.70 (0.37–0.87)	1.8 (1.4–2.7)	1.4±3.4°	2.1±3.7°	0.27
T6-7	0.7±3.2°	1.7±3.1°	0.10	0.71 (0.39–0.88)	1.7 (1.3–2.5)	0.6±2.9°	−0.2±3.1°	0.12
T7-8	−0.5±2.4°	0.1±2.5°	0.30	0.59 (0.21–0.82	1.5 (1.2–2.3)	−0.2±2.5°	−1.2±2.6°	0.11
T8-9	1.3±3.6°	0.9±2.6°	0.49	0.66 (0.31–0.85)	1.8 (1.4–2.6)	1.1±3.6°	1.5±4.1°	0.27
T9-10	4.1±3.4°	4.0±3.6°	0.83	0.83 (0.60–0.93)	1.5 (1.1–2.2)	3.0±3.1°	4.5±3.2°	0.03*
T10-11	6.5±3.1°	6.1±3.6°	0.43	0.76 (0.47–0.90)	1.7 (1.3–2.4)	6.1±4.0°	7.8±3.8°	0.03*
T11-12	7.2±2.2°	7.0±2.5°	0.64	0.69 (0.36–0.87)	1.3 (1.0–1.9)	5.8±3.2°	6.3±3.0°	0.40
T12-L1	5.0±2.4°	5.6±3.0°	0.25	0.68 (0.34–0.86)	1.5 (1.2–2.3)	5.9±3.1°	5.0±2.9°	0.06
L1-2	8.5±2.7°	8.6±3.6°	0.87	0.43 (−0.01–0.73)	2.4 (1.8–3.5)	10.2±3.4°	9.4±3.2°	0.23
L2-3	13.1±3.4°	12.5±3.3°	0.31	0.70 (0.38–0.87)	1.8 (1.4–2.7)	14.0±2.8°	13.5±3.0°	0.44
L3-4	17.3±5.7°	17.0±5.0°	0.70	0.83 (0.62–0.93)	2.2 (1.7–3.2)	16.8±4.1°	17.6±5.0°	0.22
L4-5	11.1±5.0°	10.3±4.3°	0.48	0.46 (0.02–0.75)	3.4 (2.6–5.0)	7.7±5.0°	9.8±5.7°	0.01*
L5-S1	7.9±4.6°	9.7±4.5°	0.12	0.39 (−0.06–0.71)	3.5 (2.7–5.2)	12.5±4.2°	12.0±3.6°	0.44
Means of all parameters				0.64	2.2			

*P<0.05

whether interexaminer reliability improved when the examiners used the same skin markings (see Methods section). There were significant differences between the examiners' mean values for just as many parameters as when the examiners conducted their own, independent palpation and marking of the volunteer. However, the ICCs for the two examiners' results using the same skin markings were consistently higher, ranging from 0.75 to 0.96 depending on the parameter of interest. Fifteen of the 16 parameters (94%) had an interexaminer ICC of 0.8 or higher (the remaining parameter, thoracic spine ROFE, having an ICC of 0.75); the average ICC for all parameters was 0.86 and the average SEM 3.7°.

Segmental range of flexion measures from T1-2 to L5-S1

Intraexaminer between-day reliability

For examiner 1, there were no significant differences between the mean segmental ranges of flexion measured on days 1 and 2 (Table 2). However, the corresponding ICCs were not consistently good, ranging from 0.39 to 0.83 depending on the vertebral level, with a mean value for all levels of 0.64. Only one of 17 variables had an ICC greater than 0.80, and just three had ICCs from 0.70 to

0.79. The SEMs for the various vertebral levels ranged from 1.3° (T11-12) to 3.5° (L5-S1), with an average value for all vertebral levels of 2.2°.

For examiner 2, significant differences in the mean segmental ROF were found for four of the 19 vertebral levels measured (Table 2). The ICCs ranged from 0.46 to 0.90, depending on the vertebral level examined, with a mean value for all levels of 0.67. No level had an ICC for the ROF greater than 0.80, and just four (of 17) had ICCs from 0.70 to 0.79. The SEMs ranged from 1.2° (T8-9) to 2.9° (T1-2 and T2-3), with an average value for all levels of 2.1°.

Interexaminer reliability – independent palpation and marking by each examiner

In comparing the mean values of examiner 1 and 2 on each of the two days, significant differences were observed for about a quarter of the vertebral levels examined (Table 2) (Fig. 3). The ICCs for the interexaminer measurements ranged from 0.28 to 0.81 (average 0.55) on day 1 and from 0.46 to 0.77 (average 0.62) on day 2. The SEMs for the interexaminer segmental ranges of flexion were between 1.7° and 4.0° on day 1 and between 1.6° and 3.4° on day 2, with average values for each of the two days of 2.5° and 2.3°, respectively.

flexion, *ROE* range of extension, *ROFE* range of flexion and extension

| | | Interexaminer reliability | | | | | |
| | | Day 1 | | | Day 2 | | |
ICC, day 1 vs day 2 (95% CI)	SEM (95% CI)	P (exam 1 vs 2)	ICC, exam 1 vs 2 (95% CI)	SEM exam 1 vs 2 (95% CI)	P (exam 1 vs 2)	ICC exam 1 vs 2 (95%CI)	SEM exam 1 vs 2 (95% CI)
0.72 (0.40–0.88)	2.9 (2.2–4.3)	0.00*	0.55 (0.14–0.80)	4.0 (3.0–5.8)	0.95	0.58 (0.19–0.81)	3.4 (2.6–5.0)
0.46 (0.02–0.75)	2.9 (2.9–4.2)	0.31	0.42 (–0.03–0.73)	2.9 (2.2–4.3)	0.01*	0.70 (0.38–0.87)	2.5 (1.9–3.7)
0.71 (0.39–0.88)	2.4 (1.8–3.5)	0.55	0.81 (0.57–0.92)	1.8 (1.4–2.7)	0.11	0.50 (0.07–0.77)	2.9 (2.2–4.2)
0.58 (0.19–0.81)	2.3 (1.8–3.4)	0.14	0.64 (0.28–0.84)	2.3 (1.7–3.3)	0.91	0.72 (0.41–0.88)	2.1 (1.6–3.0)
0.74 (0.45–0.89)	1.8 (1.4–2.6)	0.17	0.52 (0.11–0.78)	2.2 (1.7–3.2)	0.66	0.77 (0.50–0.91)	1.8 (1.3–2.6)
0.75 (0.46–0.89)	1.5 (1.1–2.2)	0.86	0.67 (0.32–0.86)	1.8 (1.4–2.6)	0.01*	0.64 (0.28–0.85)	1.8 (1.4–2.7)
0.46 (0.02–0.75)	1.9 (1.4–2.7)	0.68	0.28 (–0.18–0.65)	2.1 (1.6–3.0)	0.04*	0.51 (0.09–0.78)	1.8 (1.4–2.6)
0.90 (0.76–0.96)	1.2 (0.9–1.8)	0.69	0.70 (0.37–0.87)	2.0 (1.5–2.9)	0.41	0.58 (0.19–0.81)	2.2 (1.7–3.3)
0.63 (0.26–0.84)	1.9 (1.5–2.8)	0.11	0.59 (0.19–0.82)	2.1 (1.6–3.1)	0.48	0.63 (0.26–0.84)	2.1 (1.6–3.1)
0.66 (0.30–0.85)	2.3 (1.7–3.3)	0.49	0.71 (0.39–0.88)	1.9 (1.5–2.8)	0.03*	0.62 (0.24–0.83)	2.3 (1.7–3.3)
0.61 (0.23–0.83)	1.9 (1.5-2.8)	0.04*	0.48 (0.05–0.76)	2.0 (1.5–2.9)	0.22	0.69 (0.36–0.87)	1.6 (1.2–2.3)
0.73 (0.43–0.89)	1.5 (1.2–2.2)	0.13	0.56 (0.15–0.80)	1.8 (1.4–2.7)	0.33	0.55 (0.14–0.80)	2.0 (1.5–2.9)
0.59 (0.20–0.82)	2.1 (1.6–3.1)	0.04*	0.37 (–0.08–0.70)	2.4 (1.9–3.6)	0.30	0.57 (0.16–0.81)	2.2 (1.7–3.3)
0.47 (0.03–0.75)	2.1 (1.6–3.1)	0.11	0.70 (0.37–0.87)	1.7 (1.3–2.5)	0.20	0.46 (0.02–0.75)	2.3 (1.8–3.4)
0.80 (0.55–0.92)	2.1 (1.6–3.0)	0.62	0.51 (0.08–0.78)	3.5 (2.7–5.1)	0.49	0.76 (0.47–0.90)	2.5 (1.9–3.6)
0.79 (0.53–0.9)	2.5 (1.9–3.6)	0.01*	0.48 (0.05–0.76)	3.6 (2.8–5.3)	0.63	0.62 (0.25–0.83)	3.1 (2.4–4.5)
0.77 (0.49–0.90)	1.9 (1.5–2.8)	0.00*	0.34 (–0.12–0.68)	3.6 (2.7–5.3)	0.01*	0.68 (0.34–0.86)	2.3 (1.8–3.4)
0.67	2.1		0.55	2.5		0.62	2.3

Fig. 3 Comparison of segmental ranges of flexion (in degrees) for each of the two examiners on each of the 2 test days
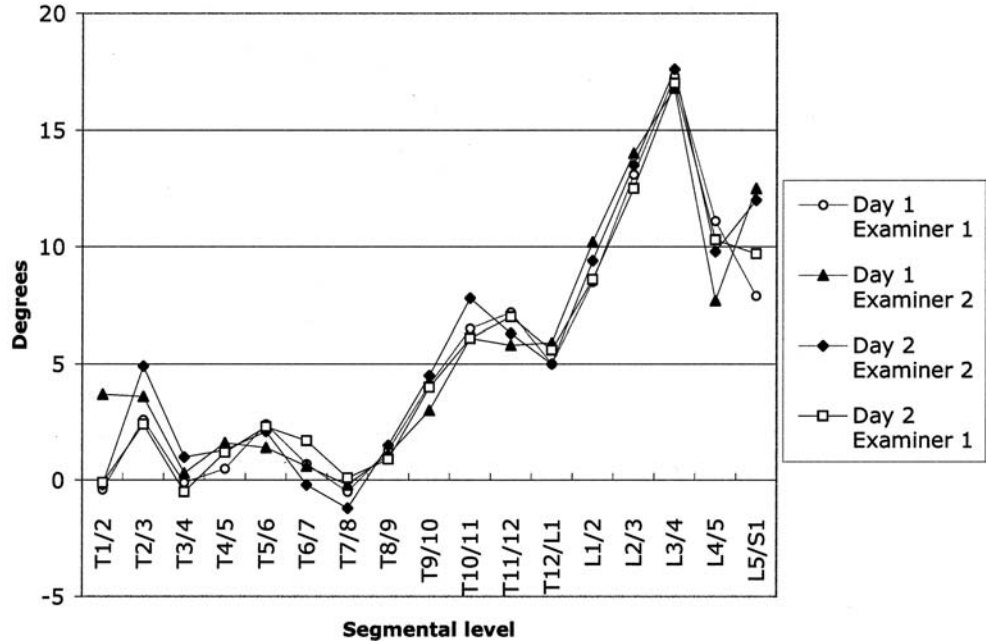
Interexaminer reliability – same skin-markings used by both examiners, day 1

When both examiners used the same skin markings for measurement, there were significant differences between their mean values for just as many parameters as when they conducted their own, independent palpation and marking of the volunteer. Although the ICCs were in general slightly higher, the average ICC for all vertebral levels was still only 0.61 (range 0.36 to 0.86, depending on level). The

Table 3 Comparison of values for range of motion measured in the present study with those reported in the literature using other devices and using X-ray measurements. All values are given in degrees. *ROF* range of flexion, *ROE* range of extension, *ROFE* range of flexion and extension, *stand* angle in standing. Not all studies examined every parameter, hence the missing values for some parameters

Device used, reference	Region measured	Lumbar lordosis	Lumbar ROF	Lumbar ROE	Lumbar ROFE	Hip stand	Hip ROF	Hip ROE
Present study[a]	T12-S1	32	65	14	79	19	48	19
Isotrak [8] [b]	L1-S1	31	55	31	90	18	75	18
Fastrak [25] [c]	L1-S1	30	54			16	72	
Fastrak [27]	L1-S1		56	20	76			
Inclinometers [32] [d]	T12-S1	23–28	42–65	12–29				
Inclinometers [52] [d]	T12-S1		42–65	14–32				
CA6000 [11]	T12-S1		81	35	116			
CA6000 [13] [e]	T12-S1		63	24	87			
CA6000 [27]	L1-S1		65	2	88			
CA6000 [29] [f]	T12-S1		57	24	81			
CA6000 [40]	T12-S1		62	21	83			
CA6000 [45] [g]	T12-S1		40–72	6–29				
X-ray [12]	L1-S1		–	–	77			
X-ray [20] [h]	L1-S1				53			
X-ray [34] [i]	L1-S1		52	16	68			
X-ray [40]	T12-S1		61	20	81			

[a]Mean, both investigators over both days
[b]Mostly females (18–40 years)
[c]Unpublished data (using the Fastrak system) from 103 healthy subjects (38% women, 62% men, age 19–59 years) involved in the first author's previous studies. Measurements were made between sensors placed over the spinous process of T1 and L1 (for the thoracic spine) and L1 and S1 (for the lumbar spine)

[d]Reviews of the results of various studies
[e]Averages over all age groups examined, both genders
[f]Averages from the 20–60-year age groups
[g]Range of mean values from the oldest to the youngest age group examined (16–90 years old)
[h]Men only (19–59 years old)
[i]Men only (25–36 years old)

ICC was greater than 0.8 for only two of 17 vertebral levels and between 0.7 and 0.8 for a further three vertebral levels. The average SEM for all levels (2.3°) was similar to that obtained when the examiners independently palpated and marked the skin (2.4°).

Discussion

General observations

The present study was carried out to examine the reliability of measures of spinal curvature and range of motion, made with the Spinal Mouse, in a group of men and women of differing age, build, habitual activity level, and occupation. All were hospital employees but carried out very different types of work (manual, technical, office, clinical). None of them had pain at the time of testing or in the preceding 2 weeks, although some had previously suffered back pain. This was therefore considered to be a sufficiently nonselective, representative sample of individuals.

Reliability of the global measurements

Between-day measurements

In the assessment of human performance capacity, 'one-off' measurements rarely provide sufficiently accurate data. In the present study, we therefore chose to carry out three trials within any given testing session and to use only the best two for further analysis. Between-day reliability was thus based on the mean values from the best two trials in any given session; the best two trials showed sufficiently consistent results to justify their averaging (data not shown due to space limitations but similar to those shown in Table 1 for between-day measures).

For the majority of parameters investigated, the ICCs were greater than 0.8, indicating good reliability [7]. The corresponding average SEM, or typical errors associated with repeated measurements, ranged between 3.5° and 4.2°. The between-day SEM can be used to indicate the degree of change required in a given individual's measures, in order to establish it (with a given level of confidence) as being a real change over and above measurement error. At the 95% confidence level, this is defined as $1.96 \times \sqrt{2} \times$ SEM or $2.77 \times$ SEM and sometimes referred to as the 'minimum (or smallest) detectable change' [3]. For lumbar ROF, for example, the minimum detectable change

was 2.77×3.2=8.9° for examiner 1 (8.3° for examiner 2). In other words, if an individual recorded a change of more than 8–9° after a given intervention, then the odds are 19:1 (i.e. 95% confidence level) that this represents a real change. The minimum detectable change can also be expressed in relation to the mean value for the given variable, to indicate the minimum detectable *percentage* change; using the same figures as above for ROF of the lumbar spine (with a mean value of approximately 64°), the minimum percentage change would equate to around 14%. Previous studies report corresponding values ranging between 5.5% [32] and 34% [17].

Some authors have actually argued that 95% confidence limits are too stringent to use as a threshold for deciding that real change has occurred, and they recommend 1.5 or 2.0 times the SEM (rather than 2.77×SEM) [21]. In this case, the corresponding odds of measuring a real change would still be 6 to 1 and 12 to 1, respectively. In the example above, this would mean that with a measured change of approximately 6° in lumbar ROF, there would be a 12 to 1 chance (92% confidence level) that this represented a real change. It is important to realise that these values for minimum detectable change per se do not necessarily indicate clinically important change; the latter is best examined in relation to changes in other clinical outcome measures. Our current preliminary studies on patients with herniated disc undergoing decompression surgery indicate that a 4-point decrease on the Roland Morris disability scale (the approximate minimum detectable change for that questionnaire [3]) is associated with an approximately 11° increase in the range of lumbar flexion [26]. Further, the difference in the improvement in lumbar flexion between a group of patients satisfied with the results of the operation and a dissatisfied group was approximately 16° [26]. It would therefore seem that the 'clinically relevant changes' reported to date using the same measurement device well exceed the error of measurement. Using the data in Table 2, Table 3, and the formulae given above, the reader can calculate the minimum detectable change for other parameters that may be of interest in specific clinical studies. Notably, the SEMs for all the measurements made in standing were particularly low, indicating that the Spinal Mouse represents a sensitive instrument for investigating changes/anomalies in the standing sagittal profile of the spine, for example associated with Scheuermann's kyphosis, flat-back syndrome, postural insufficiency, etc.

Interexaminer measurements

For some of the parameters, there were slight systematic differences in the results of the two examiners, even though their corresponding intraclass correlation coefficients were high and SEMs low. This would tend to suggest that, if implementing these measures during longitudinal studies, best results would be obtained by using just one examiner for the repeated measures made on any given individual.

However, if this were not practicable, it apparently would not have a major influence on the absolute values recorded; the interexaminer SEM was comparable to the intraexaminer SEM.

The slight systematic errors between the examiners (both reasonably experienced using the Spinal Mouse) could have arisen for a number of reasons. Firstly, there may have been differences between the two measuring devices that each experimenter used. This was considered unlikely, as the two devices were cross-checked against a fixed structure of known angle before and after the experiments, and in each case they yielded negligible differences in the angles recorded. Furthermore, if the systematic error were related to the devices themselves, then the discrepant results should have been evident for all parameters measured, which they were not – most commonly, they included the inclination angles in standing (for the hips and the trunk, generally both very low absolute values) and some of the measures for the flexion movements. A second possible explanation for the interexaminer differences could be that different start and end points were used by the two examiners during their measurements, i.e. different landmarks were palpated at the beginning. This is one of the most common sources of interexaminer error in measurements of spinal mobility [28]. However, this also appeared not to be the case here, as the small systematic differences were still evident when both examiners used the same skin markings: the intraclass correlation coefficients were somewhat higher, but the differences in mean values were similar.

Other feasible explanations for the interexaminer differences include discrepancies in the method of measurement in terms of speed, pressure exerted, and exact path followed during the rolling of the mouse. Further, depending on the sensitivity of the volunteer to the device on his/her back, differences in these factors could have slightly influenced the precise posture adopted during measurement. The recommended speed of measurement was not explicitly stated before the experiments began (and, indeed, is not specified by the manufacturers): the examiners simply carried out the tests at speeds with which they were comfortable and as they had been trained. Nonetheless, as long as the Mouse is not rolled so quickly that a signal transmission failure occurs, the speed of movement should not influence the final results. The number of data samples recorded is determined per mm distance rolled such that the speed of rolling would not effect the number of data points that contribute to the final calculated values.

As skin is flexible, the pressure exerted by the examiner in rolling the mouse along the back may influence the values recorded: greater pressure would result in the mouse traversing an apparently greater distance along the back surface and may also result in different curvatures being monitored. Examination of the lengths measured in the various postures, however, revealed no consistent differences between the examiners that could have explained

the systematic differences in angles measured. The exact path followed along the spine, i.e. whether slightly paravertebral or strictly down the midline of the spine, could contribute to slight interexaminer errors; this was not assessed in the present study and is indeed difficult to investigate. Finally, it is possible that the two examiners gave slightly different instructions to the subjects regarding the postures to be adopted, differed with respect to the stringency with which correct positioning was enforced, or encouraged the subjects to differing extents when attempting the maximum for each movement. Notably, the amount of trunk and lumbar flexion achieved by the subjects was slightly less with examiner 1 than with examiner 2. However – in common with any such voluntary test procedures – all these aspects relate more to the establishment of strict and standardised testing procedures than to the reliability of the device per se.

Validity of the global measurements

In deciding whether a new technology can be implemented for use in either the clinical or research environment, the issue of reliability is not the only factor to consider. It is also essential to confirm that the device yields valid – or accurate – results, i.e. that it really measures what it purports to measure. This is particularly important if absolute values are to be used, for example in assessing the extent of an individual's injury, severity of disease or deformity, need for rehabilitation, job suitability, etc. In relation to the current investigation, the concern is whether the angular changes recorded at the skin surface accurately represent the position and movements of the underlying vertebrae. In order to make such assessments of accuracy, an established and reliable gold standard is required with which the new system can be directly compared in a given group of individuals using, for example, the method of Bland and Altman [4]. With regard to the accuracy of measures of lumbar spinal mobility, this is difficult to do, because no suitable gold standard currently exists.

Although radiographic measures of range of motion are commonly considered to represent the ideal, no study has ever demonstrated acceptable reliability for X-ray measures of spinal range of motion. A number of studies have examined the error associated with repeated measurement of vertebral angles on a given set of radiographic films and found that it can sometimes be appreciable [28, 33, 39, 50]. However, this represents only one likely source of error associated with the whole procedure: in addition to the interpretation and measurement of the final X-ray films, measurement error can also arise as a result of differences in patient positioning during imaging, image quality, patient performance, and so on. Thus, the bottom line is that if the X-ray technique itself has not been shown to be reliable, then its application can hardly

be considered to be a gold standard. Under these circumstances, the best that can be done is to ensure approximate agreement (convergent validity) between the mean values measured with the Spinal Mouse and those measured with various other devices, including radiological images. This should at least indicate whether the device is yielding relatively realistic values.

In view of (1) the fact that skin-surface measurements by definition follow the line of the posterior elements and not that of the vertebral bodies (as in X-ray measures) and (2) the varying distribution of subcutaneous tissue overlying the spine, most notably in the lumbar region towards the sacrum, it would not necessarily be expected that the absolute curvatures measured with the Spinal Mouse would be directly comparable to those measured radiographically. They should, however, be comparable to those measured using other skin-surface devices. Furthermore, the ranges of motion measured for the various regions of the spine should be comparable for both skin-surface devices and X-ray measurements, assuming the same number of motion segments is included in the measurements.

Comparison of Spinal Mouse
measures of global spinal curvature and range of motion with other methods of measurement

Thoracic spine

A number of radiographic studies, but few studies using skin-surface devices, have documented normal values for the sagittal curvature of the thoracic spine. The standing kyphosis measured in the present study with the Spinal Mouse was on average 45°; this compares favourably with our previous, unpublished data using the Fastrak system [25], in which the average value from a total of 103 healthy subjects (38% women and 62% men aged 19–59 years) was 43±13°. It also compares well with published radiographic measures, e.g. 48° [16], 46° [18], 42° [22], 47° [49], and 34° [50]. As there is typically only minimal subcutaneous tissue in this region of the spine, similar values for skin-surface and radiographic measures may indeed be expected.

Thoracic flexion and extension are considered to be relatively limited motions [2], and few studies have reported normal values for these. For the 103 subjects in the study described above [25], the mean value for the thoracic ROF was 34±14°; this compares reasonably well with the average value of 25° recorded in the present study. These values are not dissimilar from that (30°) reported for the normal thoracic ROF in clinical handbooks [24]. In both the current and previous studies [25] using skin-surface devices, the mean range of extension was almost negligible. This may be because, in standing, full extension of the thoracic spine is difficult to achieve under the constraints necessary for measurement (head/neck in a

neutral position and the hands crossed over the chest). The few previously available reports in the literature refer to an average range of extension for the thoracic spine of approximately 20° [24].

Lumbar spine and hips

A large number of investigations have examined the curvature and range of motion of the lumbar spine and hips, perhaps reflecting the somewhat greater clinical interest in these regions of the spine (Table 3). Many of the skin-surface devices effectively zero the device when the subject is in upright standing (i.e. the angle of the lordosis in upright standing is set to zero and the spinal movements are measured relative to this starting point), and hence they are unable to record an *absolute* value for standing hip and lumbar curvatures. The standing hip angle has been reported in only a few other studies using skin-surface devices, and the values measured in the present study compare favourably with these (all approximately 16–19°) (Table 3). Lumbar lordosis, as measured with inclinometers, has been reported in a number of studies, and the values vary between 23° and 33° [32]. These values are similar to those recorded with the Isotrak or Fastrak (30° and 31°, respectively) (Table 3) and in the present study (32°). All the skin-surface devices tend to yield considerably lower values for standing lumbar lordosis than those measured with X-ray (for T12-S1 between 47° and 64°) [16, 19, 22, 30, 47, 49, 50, 52], most likely for the reasons explained earlier (i.e. the curvature is not that formed by the bony structures themselves, and there can sometimes be thicker layers of subcutaneous tissue overlying the lower lumbar spine and sacrum).

The mean ROF of the lumbar spine, as measured with the Spinal Mouse, lies well within the range previously reported in the literature (Table 3). Naturally some caution is required in making these comparisons, as the gender, age, and back pain history of the participants all differ somewhat between the studies, as does the number of motion segments attributed to the lumbar spine. With respect to the latter, most of the skin-surface devices – CA6000 [11, 13, 29, 37, 41, 46], long-arm goniometers [40], and inclinometers [53] – measure from the thoracolumbar junction (T12-L1) to the sacrum, and these obviously give slightly higher values than when the motion sensors are placed over the spinous process of L1 and the sacrum, as is typically done when using the Fastrak [25, 27] or Isotrak [8, 9, 10] systems or when the curvature is measured from L1-2 to L5-S1 with X-rays [12, 34]. Regarding spinal mobility, there appears to be no consensus as to whether the T12-L1 motion segment should be considered part of the lumbar spine, although guidelines for the measurement of lumbar ROM [2] recommend its inclusion, as T11-12 is considered to represent the functional junction between the thoracic and lumbar regions [5].

The values for lumbar spine range of extension, as measured with the Spinal Mouse, appear to be somewhat lower than those reported using other skin-surface devices, although they are relatively close to those determined using X-ray analysis (Table 3). In our experience with devices that are affixed to the skin, there is sometimes a tendency for the sensors to tilt backwards under gravity or for the skin to buckle in standing extension such that the sensors do not always remain parallel to the skin surface (in a manner which exaggerates the curvature of the extended spine). Alternatively, the standing position adopted during the present study may not have allowed the achievement of maximal extension, as discussed earlier in relation to the measurements of thoracic spine extension. Extension is considered to be a notoriously difficult posture to measure using skin-surface devices [32].
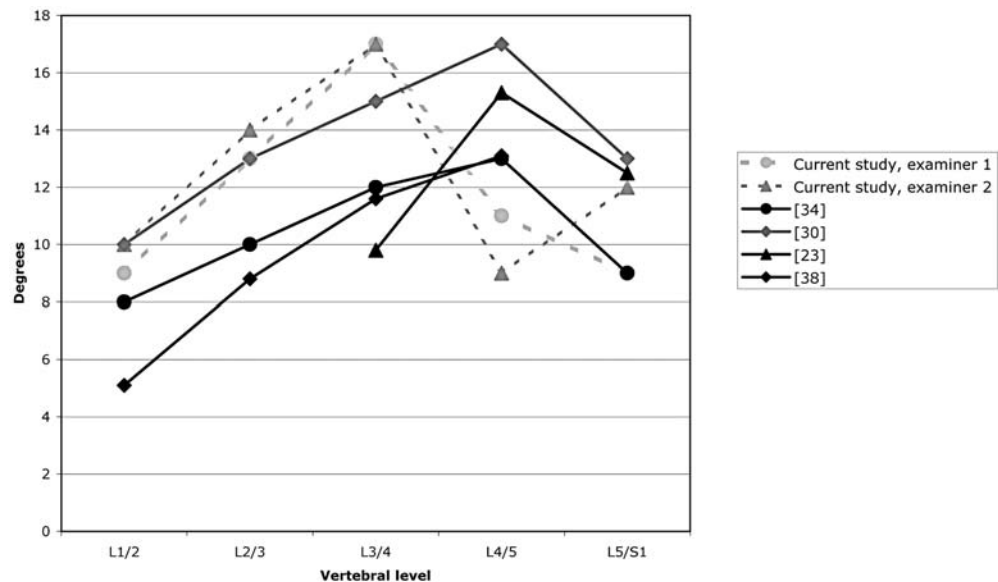
Reliability of the lumbar segmental measurements

As might be expected, in view of the smaller angles being measured, neither the intraexaminer nor interexaminer reliability was as good for segmental mobility as for the global spinal segments. The intraclass correlation coefficients (ICCs) were high for certain segments, but overall the reliability for the segmental angles was rather low, with an average ICC of just 0.6 over all segments. The SEMs were on average approximately 2–3°, giving a smallest detectable change of 6–8° (95% confidence level) or 4–5° (92% confidence level).

We consider this magnitude of error, although similar to that reported for radiographic measures of segmental range of motion [15, 33, 44], too high to allow reliable interpretation of individual values and their changes over time, at least in normal volunteers. Indeed, for some segments, the error of measurement was greater than the mean value itself. Nonetheless, it is possible that in patients with spinal instability and thus abnormally high ranges of motion for individual segments, the Spinal Mouse may be sensitive enough to detect such changes. This will be investigated in our future studies.

The group mean values showed minimal differences between days or between investigators (Fig. 3), which would tend to suggest that, for a given group of individuals, the mean values may still be of use in interpreting trends for change, for example in response to a given pathology or following surgical intervention. Some of our preliminary (unpublished) clinical studies have shown mean changes of 4–6° in the segmental ROF of operated or neighbouring segments after spinal surgery, suggesting that the device may indeed be sensitive enough to detect relative change in segmental mobility in specific groups of patients.

134

Fig. 4 Comparison of the segmental ranges of flexion measured in the current study with those reported in the literature

Validity of the lumbar segmental measurements

The concurrent validity of lumbar segmental ranges of flexion measured in the current study is difficult to assess, as few skin-surface devices have been used to examine segmental angles. Indeed, such small angles are difficult to measure accurately with the motion analysis devices typically used for measuring global movements, as the sensors themselves can be almost as long as the intervertebral distances. Earlier attempts to measure segmental motion from surface changes in back curvature proved to be unreliable and inaccurate [43], although measures of regional (i.e. upper vs lower lumbar) mobility were more successful [5].

The mean values for the segmental ranges of flexion recorded with the Spinal Mouse, and in particular the relative differences in segmental mobility along the length of the lumbar spine, compare reasonably well with those measured using X-ray films (Fig. 4), with the exception of the values recorded at L4-5 and L5-S1. In most X-ray studies, these two lower segments of the lumbar spine demonstrate the greatest ROF, but with the Spinal Mouse the greatest mobility was recorded for L3-4. Interestingly, in the present study, the greatest measurement error was frequently recorded for the lowest two segments, L4-5 and L5-S1. This was also reported to be the case for radiographic measures [15, 22, 44]. Further, in a previous comparison of radiographic and surface measurements, the lowest correlations between the two procedures were found for the segments L4-5 and L5-S1 [43]. If both skin-surface and radiographic techniques are subject to greatest error at these lower lumbar levels, it is difficult to ascertain which method yields more accurate results. It is unfortunate that this very same region is the one most fre-

quently affected in common disorders of the lumbar spine and therefore represents the site of greatest clinical interest. Perhaps the performance of some simple Spinal Mouse measurements as a complementary test procedure in more invasive studies of segmental motion (in which motion sensors are attached to pins or wires fixed in the spinous processes) may shed light on the relative accuracy of radiographic and skin-surface measurements.

Conclusion

For global regions of the spine, the Spinal Mouse delivered consistently reliable results for standing curvatures and ranges of motion both within and between days and also between investigators. The values compared well with those previously reported in the literature. This suggests that the device can be used with confidence in both research and clinical environments for the measurement of sagittal profile and/or range of motion of global regions of the spine. It may find clinical application in the assessment of structural deformity associated with, for example, Scheuermann's disease, osteoporotic kyphosis, scoliosis, or flat-back syndrome or in the monitoring of disturbances/restrictions in movement in connection with 'mechanical' spinal disorders such as herniated disc, simple mechanical back pain, spinal instability, etc. Further, in the areas of ergonomics, work place, and seating design, the device may be of use in assessing the spinal curvature associated with postures commonly adopted during the performance of given tasks.

As might be expected for the smaller angles being measured, the segmental ranges of flexion measured with the Spinal Mouse showed less reliability, as judged by the

lower ICCs. Their usefulness for the interpretation of change on an individual basis thus remains questionable. Nonetheless, the average SEM for between-day measurements was relatively low (approximately 2–3°), and group mean values showed little between-day differences, suggesting that the examination of mean segmental ranges of flexion, for groups of individuals with a given spinal pathology or undergoing some type of intervention, may still be of clinical relevance.

References

1. Adams MA, Dolan P (1991) A technique for quantifying the bending moment acting on the lumbar spine in vivo. J Biomech 24:117–126
2. American Medical Association (2000) The spine. In: Cocchiarella L, Andersson GBJ (eds) Guides to the evaluation of permanent impairment. Fifth edn. AMA, Chicago
3. Beaton DE (2000) Understanding the relevance of measured change through studies of responsiveness. Spine 25:3192–3199
4. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, Feb 8:307–310
5. Burton AK (1986) Regional lumbar sagittal mobility; measurement by flexicurves. Clin Biomech 1:20–26
6. Chinn S (1991). Repeatability and method comparison. Thorax 46:454–456
7. Currier DP (1990) Elements of research in physical therapy. Third edn. Williams and Wilkins, Baltimore
8. Dolan P, Adams MA (1993) Influence of lumbar and hip mobility on the bending stresses acting on the lumbar spine. Clin Biomech 8:185–192
9. Dolan P, Earley M, Adams MA (1994) Bending and compressive stresses acting on the lumbar spine during lifting activities. J Biomech 27:1237–1248
10. Dolan P, Greenfield K, Nelson RJ, Nelson IW (2000) Can exercise therapy improve the outcome of microdiscectomy? Spine 25:1523–1532
11. Dopf CA, Mandel SS, Geiger DF, Mayer PJ (1994) Analysis of spine motion variability using a computerized goniometer compared to physical examination. A prospective clinical study. Spine 19:586–595
12. Dvorak J, Panjabi MM, Chang DG, Theiler R, Grob D (1991) Functional radiographic diagnosis of the lumbar spine. Flexion-extension and lateral bending. Spine 16:562–571
13. Dvorak J, Vajda EG, Grob D, Panjabi MM (1995) Normal motion of the lumbar spine as related to age and gender. Eur Spine J 4:18–23
14. Essendrop M, Maul I, Laubli T, Riihimaki H, Schibye B (2002) Measures of low back function: a review of reproducibility studies. Clin Biomech 17:235–249
15. Frobin W, Brinckmann P, Leivseth G, Biggemann M, Reikeras O (1996) Precision measurement of segmental motion from flexion-extension radiographs of the lumbar spine. Clin Biomech 11:457–465
16. Gelb DE, Lenke LG, Bridwell KH, Blanke K, McEnery KW (1995) An analysis of sagittal spinal alignment in 100 asymptomatic middle and older aged volunteers. Spine 20:1351–1358
17. Gill K, Krag MH, Johnson GB, Haugh LD, Pope MH. (1988) Repeatability of four clinical methods for assessment of lumbar spinal motion. Spine 13:50–53
18. Harrison DE, Cailliet R, Harrison DD, Janik TJ (2002) How do anterior/posterior translations of the thoracic cage affect the sagittal lumbar spine, pelvic tilt, and thoracic kyphosis? Eur Spine J 11:287–293
19. Harrison DE, Harrison DD, Cailiet R, Janik TJ, Holland B (2001) Radiographic analysis of lumbar lordosis: centroid, Cobb, Trall, and Harrison posterior tangent methods. Spine 26:235–242
20. Hayes MA, Howard TC, Gruel CR, Kopta JA (1989) Roentgenographic evaluation of lumbar spine flexion-extension in asymptomatic individuals. Spine 14:327–331
21. Hopkins WG (2000). Measures of reliability in sports medicine and science. Sports Med 30:1–15
22. Jackson RP, McManus AC (1994) Radiographic analysis of sagittal plane alignment and balance in standing volunteers and patients with low back pain matched for age, sex, and size. Spine 19:1611–1618
23. Kanayama M, Abumi K, Kaneda K, Tadano S, Ukai T (1996) Phase lag of the intersegmental motion in flexion-extension of the lumbar and lumbosacral spine. An in vivo study. Spine 21:1416–1422
24. Louis R (1983) Chirurgie du rachis. Springer, Berlin Heidelberg New York, pp 64–74
25. Mannion AF, Dolan P, Dumas GA, Stevenson J, Cooper RG, Wood K, Adams MA (unpublished data) pooled results from 3 previously published studies, in which part of the investigation also required measurements of spinal mobility with the Fastrak system.
26. Mannion AF, Dvorak J, Müntener M, Grob D (2003). Prospective study of changes in self-rated disability and range of motion of the lumbar spine in patients with disc herniation undergoing decompression surgery. Presented at the Annual Meeting of the International Society for the Study of the Lumbar Spine, Vancouver
27. Mannion AF, Troke M (1999) A comparison of two motion analysis devices used in the measurement of lumbar spinal mobility. Clin Biomech 14:612–619
28. Mayer RS, Chen IH, Lavender SA, Trafimow JH, Andersson GBJ (1995) Variance in the measurement of sagittal lumbar spine range of motion among examiners, subjects, and instruments. Spine 20:1489–1493
29. McGregor AH, McCarthy ID, Hughes SP (1995) Motion characteristics of the lumbar spine in the normal population. Spine 20:2421–2428
30. Miyasaka K, Ohmori K, Suzuki K (2000) Radiographic analysis of lumbar motion in relation to lumbosacral stability. Spine 25:732–737
31. Morin Doody M, Lonstein JE, Stovall M, Hacker DG, Luckyanov N, Land CE (2000) Breast cancer mortality after diagnostic radiography: findings from the U.S. Scoliosis Cohort Study. Spine 25:2052–2063
32. Ng K-F, Kippers V, Richardson CA, Parnianpour M (2001) Range of motion and lordosis of the lumbar spine. Reliability of measurement and normative values. Spine 26:53–60
33. Panjabi M, Chang D, Dvorak J (1992) An analysis of errors in kinematic parameters associated with in vivo functional radiographs. Spine 17:200–205
34. Pearcy M, Portek I, Shephard J (1984). Three dimensional X-ray analysis of normal movement in the lumbar spine. Spine 9:294–297

35. Pearcy MJ, Hindle RJ (1989) New method for the non-invasive three-dimensional measurement of human back movement. Clinical Biomechanics 4:73–79
36. Perneger TV (1998) What's wrong with Bonferroni adjustments? BMJ 316:1236–1238
37. Peterson CM, Johnson RD, Schuit D, Hayes KW (1994) Intraobserver and interobserver reliability of asymptomatic subjects thoracolumbar range of motion using the OSI CA6000 spine motion analyser. J Orthop Sports Phys Ther 20:207–212
38. Plamondon A, Gagnon M, Maurais G (1988) Application of a stereoradiographic method for the study of intervertebral motion. Spine 13:1027–1032
39. Polly DW Jr, Kilkelly FX, McHale KA, Asplund LM, Mulligan M, Chang AS (1996) Measurement of lumbar lordosis. Evaluation of intraobserver, interobserver, and technique variability. Spine 21:1530–1535
40. Salisbury PJ, Porter RW (1986) Measurement of lumbar sagittal mobility. A comparison of methods. Spine 12: 190–193
41. Schuit D, Petersen C, Johnson R, Levine P, Knecht H, Goldberg D (1997) Validity and reliability of measures obtained from the OSI CA-6000 Spine Motion Analyzer for lumbar spinal motion. Manual Therapy 2:206–215
42. Seichert N, Baumann M, Senn E, Zuckriegl H (1994) Die Rückenmaus – ein analog-digitales Messgerät zur Erfassung der sagittalen Rückenkontur. Phys Rehab Kur Med 4:34–43
43. Stokes IAF, Bevins TM, Lunn RA (1987) Back surface curvature and measurement of lumbar spinal motion. Spine 12:355–361
44. Tallroth K, Ylikoski M, Landtman M, Santavirta S (1994) Reliability of radiographical measurements of spondylolisthesis and extension-flexion radiographs of the lumbar spine. Eur J Radiol 18:227–231
45. Tillotson KM, Burton AK (1991) Non-invasive measurement of lumbar sagittal mobility. An assessment of the flexicurve technique. Spine 16:29–33
46. Troke M, Moore AP, Maillardet FJ, Hough A, Cheek E (2001) A new, comprehensive normative database of lumbar spine ranges of motion. Clin Rehabil 15:371–379
47. Troyanovich SJ, Cailliet R, Janik TJ, Harrison DD, Harrison DE (1997) Radiographic mensuration characteristics of the sagittal lumbar spine from a normal population with a method to synthesize prior studies of lordosis. J Spinal Disord 10:380–386
48. Van Herp G, Rowe P, Salter P, Paul JP (2000) Three-dimensional lumbar spinal kinematics: a study of range of movement in 100 healthy subjects aged 20 to 60+ years. Rheumatology (Oxford) 39:1337–1340
49. Vaz G, Roussouly P, Berthonnaud E, Dimnet J (2002) Sagittal morphology and equilibrium of pelvis and spine. Spine 11:80–87
50. Vedantam R, Lenke L, Keeney J, Bridwell K (1998) Comparison of standing sagittal spinal alignment in asymptomatic adolescents and adults. Spine 23:211–215
51. Vogt L, Pfeifer K, Portscher M, Banzer W (2001) Influences of nonspecific low back pain on three-dimensional lumbar spine kinematics in locomotion. Spine 26:1910–1919
52. Voutsinas SA, MacEwen GD (1986) Sagittal profiles of the spine. Clin Orthop 210:235–242
53. Zuberbier OA, Kozlowski AJ, Hunt DG, Berkowitz J, Schultz IZ, Crook JM, Milner RA (2001) Analysis of the convergent and discriminant validity of published lumbar flexion, extension, and lateral flexion scores. Spine 26: 472–478