Journal of Biomechanics 121 (2021) 110435

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

Reliability of measures to characterize lumbar movement patterns, in repeated seated reaching, in a mixed group of participants with and without low-back pain: A test-retest, within- and between session

Meta H. Wildenbeest [a,b,*], Henri Kiers [a], Matthijs Tuijt [a], Jaap H. van Dieën [b]

[a] HU University of Applied Sciences, Institute for Human Movement Studies, Postbus 12011, 3501 AA Utrecht, the Netherlands
[b] Department of Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, 1081 BT Amsterdam, the Netherlands

ARTICLE INFO

Article history:
Accepted 9 April 2021

Keywords:
Test-retest
Spinal control
Reliability
Stability
Variability

ABSTRACT

Literature highlights the need for research on changes in lumbar movement patterns, as potential mechanisms underlying the persistence of low-back pain. Variability and local dynamic stability are frequently used to characterize movement patterns. In view of a lack of information on reliability of these measures, we determined their within- and between-session reliability in repeated seated reaching.

Thirty-six participants (21 healthy, 15 LBP) executed three trials of repeated seated reaching on two days. An optical motion capture system recorded positions of cluster markers, located on the spinous processes of S1 and T8. Movement patterns were characterized by the spatial variability (meanSD) of the lumbar Euler angles: flexion–extension, lateral bending, axial rotation, temporal variability (CyclSD) and local dynamic stability (LDE). Reliability was evaluated using intraclass correlation coefficients (ICC), coefficients of variation (CV) and Bland-Altman plots. Sufficient reliability was defined as an ICC $\geq$ 0.5 and a CV < 20%. To determine the effect of number of repetitions on reliability, analyses were performed for the first 10, 20, 30, and 40 repetitions of each time series.

MeanSD, CyclSD, and the LDE had moderate within-session reliability; meanSD: ICC = 0.60–0.73 (CV = 14–17%); CyclSD: ICC = 0.68 (CV = 17%); LDE: ICC = 0.62 (CV = 5%). Between-session reliability was somewhat lower; meanSD: ICC = 0.44–0.73 (CV = 17–19%); CyclSD: ICC = 0.45–0.56 (CV = 19–22%); LDE: ICC = 0.25–0.54 (CV = 5–6%).

MeanSD, CyclSD and the LDE are sufficiently reliable to assess lumbar movement patterns in single-session experiments, and at best sufficiently reliable in multi-session experiments. Within-session, a plateau in reliability appears to be reached at 40 repetitions for meanSD (flexion–extension), meanSD (axial-rotation) and CyclSD.

1. Introduction

Chronic low-back pain (LBP) is widely prevalent and entails large economic costs (Hartvigsen et al., 2018). Hartvigsen et al. (2018) emphasized the need for research into the mechanisms underlying the persistence of LBP, and changed lumbar motor control is assumed to be one of these mechanisms (Van Dieën et al., 2019a,b). Individuals with LBP move differently than people without pain (Van Dieën et al., 2019a,b). It is unknown why individuals with LBP persist in their altered movement behavior, after healing of the injured tissue. In addition to pain itself, pain-related factors, like impaired proprioception and fear of re-injury, are assumed to play a role (Van Dieën et al., 2019a,b). To study such mechanisms experimentally, reliable characterization of lumbar movement patterns is a prerequisite. Frequently used measures to characterize movement patterns are variability and local dynamic stability (Stergiou and Decker, 2011; Mavor and Graham, 2015; Bruijn et al., 2013).

Variability and local dynamic stability may provide information on the quality of the control of lumbar movement, and probably more so than currently widely used clinical measures, such as range of motion, muscle strength, and endurance (Dupeyron et al., 2013). Variability reflects the spatial and/or temporal variability of repeated movement cycles (cyclSD). Spatial variability is commonly expressed as the mean standard deviation of lumbar angles over cycles (meanSD; Dingwell and Marin, 2006) and

* Corresponding author at: HU University of Applied Sciences, Institute for Human Movement Studies, Postbus 12011, 3501 AA Utrecht, the Netherlands.
E-mail address: meta.wildenbeest@hu.nl (M.H. Wildenbeest).

https://doi.org/10.1016/j.jbiomech.2021.110435
0021-9290/

temporal variability as the standard deviation of cycle times (Marques et al., 2017). Local dynamic stability reflects sensitivity to small perturbations and is expressed by the local divergence exponent (LDE), which quantifies the relative rate of divergence between neighboring trajectories in a reconstructed state space of lumbar kinematics (Bruijn et al., 2013; Mehdizadeh, 2019). These measures are complementary, and inform about different aspects of motor control (Stergiou and Decker, 2011).

Several authors have reported within- and between-sessions reliability of the LDE in healthy participants during walking (van Schooten et al., 2013; Reynard and Terrier, 2014; Kang and Dingwell, 2006). These studies focused on trunk movements relative to a global reference system, whereas lumbar movement, i.e. relative movement between thorax and pelvis, may be more pertinent. Dupeyron et al., (2013) reported good within-session reliability (ICC = 0.87–0.97) of the LDE of lumbar angles during repeated reaching. However, this study assessed healthy participants only, within a single session. Studies on pain persistence are likely to require multiple sessions, and participants with and without LBP, where in the former group pain levels can be expected to be quite variable over days, potentially increasing between-sessions variance. Recently, a between-day ICC-value of 0.49 was reported for meanSD, for repeated reaching (flexion/extension task), in patients with LBP (Graham et al., 2020). In both of these studies, tests were performed while standing, while adults spend more than half their waking hours in sedentary activities (Owen et al., 2009). In addition, it's recommended that future studies will assess changes in trunk movement after exposure to trunk perturbations, for which seated testing would have the advantage of excluding effects of leg movements (Maaswinkel et al., 2016).

Therefore, the goal of the present study was to assess reliability of variability and local dynamic stability during repetitive seated reaching, over multiple test moments within and between days, in a mixed group of people with and without low back pain. Based on previous studies (Dupeyron et al., 2013; Graham et al., 2020), we hypothesized that reliability, within- and between-sessions, would be moderate but sufficient for usage of the measures studied in future research.

2. Methods

2.1. Participants

Thirty-six adults volunteered for this study: 21 healthy participants and 15 participants with LBP. Mean age was 35.8 years (SD 13.6), mean height 1.78 m (SD 0.09) and mean mass was 76.01 kg (SD 10.2). Inclusion criteria for participants with LBP were: (1) > 1 episode of non-specific low back pain, or continuous non-specific low back pain within the last 2 years; (2) duration of an episode of low back pain $\geq$ 2 weeks; (3) pain intensity affected by posture or movement. The latter was to focus on patients with pain originating from a nociceptive source. Healthy participants were included when free from episodes of non-specific low-back pain for 2 years. Exclusion criteria for both groups were (1) perceived balance problems; (2) BMI > 30 in combination with high abdominal circumference (males > 102 cm, females > 88 cm); (3) any systemic pathology (e.g. Parkinson's, diabetes mellitus or cancer), earlier spine surgery, infections, medication which might influence movement patterns (antidepressants, analgesics, tranquillizers), pregnancy, cardiovascular pathology, neurologic pathology, respiratory ailments, or significant musculoskeletal injury in the past 6 months.

Participants with LBP completed two questionnaires: the Oswestry Disability Index (ODI; Fairbanks et al., 1980), with a minimum score indicating no disability and the maximum score indicating 100% disability, and the StarT back screening tool (SBST; Hill et al., 2008), with a minimum score 0–3 indicating low risk for the presence of psychosocial prognostic factors and a score $\geq$ 4 (subscale question 4–9) indicating high risk. To evaluate the pain intensity at the time of testing, a numerical rating scale (NRS) was used (0 = no pain, 10 = most imaginable pain). Prior to participation, all participants provided informed consent. The protocol had been approved by the ethical committee of the Faculty of Human Movement Sciences, VU University of Amsterdam (VCWE-2019-013).

2.2. Materials

A custom-made chair, without back rest and arm supports, was rigidly attached to a DynSTABLE (Motek Medical Amsterdam, Netherlands) platform with dimensions: 1x1x0.3 m (Width $\times$ Depth $\times$ Height). Marker locations were recorded using a motion capture system consisting of 4 Vicon Bonita3 cameras (VICON-612 system, Oxford Metrics, UK), and sampled with D-Flow software at approximately 100 samples/s (Motek Medical Amsterdam, Netherlands). To assess lumbar motion, two clusters of three markers were used. Clusters were fixed to the spinous processes of T8 and S1using adhesive tape (Fig. 1A) (Dupeyron et al., 2013).

2.3. Experimental procedure

Participants reached forward 45 times, while seated with one arm crossed in front of the chest (Fig. 1B). Before the trial, participants practiced 5 times to get used to the task. Repeated forward reaching at preferred speed was performed from upright posture to a flexed position. The fingers of the dominant hand pressed the button of a joystick situated in front of the participant, at knee level, and at a distance of 125% of the length of the upper limb (Fig. 2). A color-marked Borg-RPE-Scale (0 to 10) was completed after each trial, to asses fatigue in the lumbar region. Participants visited the lab on two days. On day one, trials 1A and 1B were executed. Between trials, the participant got off the DynSTABLE, for a

Fig. 1A. Fixation of marker clusters on the thorax and pelvis (T8 and S1).

Fig. 1B. Seated posture during reaching task.

minimum of ten minutes. Marker clusters were not removed between trials 1A and 1B, to allow assessment of reliability within a single experimental session with multiple trials. More than one day later, participants executed trial 2 (mean 7.9 days (SD 6.4)).

2.4. Joint kinematics

The final 40 of the 45 repetitions were selected for analysis, to omit clear transients that sometimes occurred in the first cycles. Data were (cubic spline) interpolated to 100 Hz, to account for missing data, and to correct for small fluctuations in sample rate, caused by D-flow software, recording at 102/103 Hz. Segment orientations were computed in the global axis system. Subsequently, relative orientations between thorax and pelvis were determined and decomposed into lumbar angles using Euler decomposition in the order flexion/extension, lateral bending, torsion.

2.5. Variability

The time series of the lumbar angles were divided into cycles. For cycle detection (one cycle is one reaching movement), the most forward sagittal plane orientation of the thorax (T8) in each cycle, was detected using a peak detection algorithm. CyclSD was quantified as the standard deviation of the cycle durations. For spatial variability, lumbar angle data for each reaching cycle were normalized to 101 samples (0–100%) for flexion–extension, latero-flexion and torsion. Cross-correlation was used to optimally align all repetitions.

MeanSD was calculated as the average of the standard deviations at all normalized time point across the cycles (Dingwell and Marin, 2006).

2.6. Local dynamic stability

Using cubic spline interpolation, lumbar angle time series were normalized to a fixed number of data points (300 times the number of cycles), as the number of samples affects the LDE (Bruijn et al., 2009). A 6-dimensional state-space was reconstructed using the three lumbar angles, and a 30-samples (10% of the average number of samples per cycle) time-delayed copy. The revised Rosenstein method was used, to minimize effects of noise in the time series (Mehdizadeh, 2019), by tracking divergence between kinematic states evolving from each data point and its 15 nearest neighbors. Divergence curves were logarithmically transformed, and averaged over the nearest neighbors per reference point and over all reference points. The LDE was determined as the slope of the line of best fit over the first 0.25 cycle of the resulting divergence curve (Graham et al., 2012). The algorithm used for LDE calculation, is available at http://doi.org/10.5281/zenodo.573285, https://zenodo.org/record/4681213 (Bruijn, 2021).

2.7. Statistical analyses

Statistical analyses were performed with IBM SPSS Statistics 25 software. Agreement between the measurements within- and between-sessions was assessed by the Intraclass Correlation Coefficient (ICC) for all variables and by Bland-Altman plots. ICC form (2-way random, single measurement, absolute agreement) was selected according to the guidelines of Koo and Li (2016). The coefficient of variation was assessed with the method of Hyslop and White (2009). Sufficient reliability was defined as a combination of an ICC $\geq$ 0.5 and a CV < 20%. For composing the Bland & Altman plots, first mean differences and $SD_{differences}$ were calculated, next, limits of agreement (LoA) (mean difference ± 1.96 × $\sigma_{difference}$). Finally independent-t-tests were performed, to check if the mean difference was significant (De Vet et al., 2011). To determine how the number of repetitions affects reliability, these analyses were performed for the first 10, 20, 30, and 40 cycles of each time series.

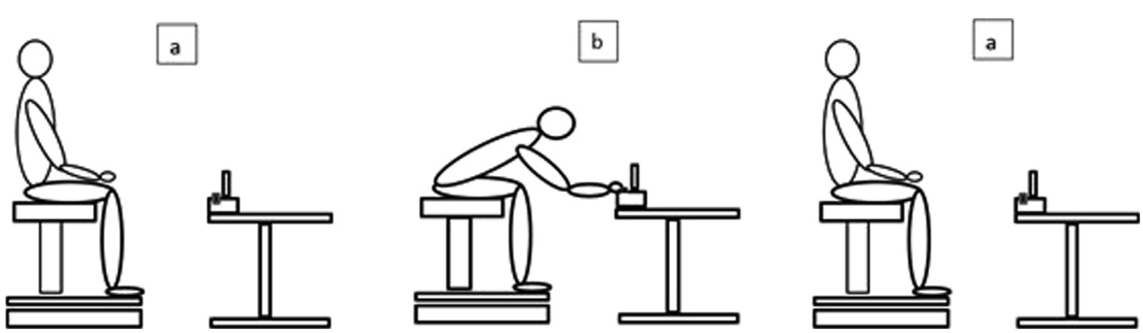

Fig. 2. Schematic overview of the task performed by the participants. Starting position – upright sitting (a), forward reaching to a distance of 1.25 × length of the upper limb to reach the target (b), and starting position again (a).

3. Results

3.1. Participants

LBP participants showed a low level of disability (mean ODI score 17.7 (±15.2)) and low risk for chronicity (mean STBST score 1.7 (±1.6)). No significant differences in pain intensity between days or perceived exertion between trials were found (Table 1).

The mean number of samples per cycle was 314.8 (SD = 50.2). In two out of 108 trials, samples were missing, because one of the markers was blocked from the camera's. In one trial 40 consecutive samples were missing and this measurement was excluded (trial 3 of 1 healthy participant). In the other trial interpolation was used to correct for short intervals < 6 samples.

3.2. Intra-class correlations

Except for cyclSD (mean difference of 0.016), there were no systematic differences between the results of trials 1A and 1B. The between-sessions differences in cyclSD and LDE were significant between trials 1A and 2 (Table 2). Within-session ICC's for meanSD were 0.71 ($\text{meanSD}_{\text{flexion-extension}}$), 0.60 ($\text{meanSD}_{\text{lateral-bending}}$) and 0.73 ($\text{meanSD}_{\text{axial-rotation}}$). Between-sessions ICC's for meanSD were 0.73 and 0.69 for $\text{meanSD}_{\text{flexion-extension}}$, 0.65 and 0.44 for $\text{meanSD}_{\text{lateral-bending}}$, 0.47 and 0.51 for $\text{meanSD}_{\text{axial-rotation}}$. The within-session ICC for cyclSD was 0.68. The between-sessions ICC's for cyclSD were 0.45 (trial 1A-2) and 0.56 (trial 1B-2). The within-session ICC for local dynamic stability was 0.62. The between-sessions ICC's for local dynamic stability were 0.25 (trial 1A-2) and 0.54 (trial 1B-2) (Table 3).

3.3. Limits of agreement

The limits of agreement for $\text{meanSD}_{\text{flexion-extension}}$ were similar to the mean values. The limits of agreement for $\text{meanSD}_{\text{axial-rotation}}$, $\text{meanSD}_{\text{lateral-bending}}$ and for cyclSD, were similar to, or even exceeded the mean values (Fig. 3 and Fig. 4). The limits of agreement for the LDE measurements were substantially smaller than the mean values (Fig. 5).

3.4. Coefficients of variation

CV, within- and between-sessions, for meanSD, for all three degrees of freedom, and for cyclSD, were approximately 17%. For the LDE, the CV was approximately 5% (Table 2).

3.5. Number of repetitions

Analyses were performed the first 10, 20, 30, 40 repetitions of each time series (Fig. 6). Within-session, the magnitude of the ICC's increased with the number of repetitions analyzed, with the exception of the $\text{meanSD}_{\text{lateral-bending}}$, and an outlier for cyclSD after 10 repetitions. The increase of ICC values decreased with a higher number of repetitions until for most variables a plateau was reached after 30 repetitions. However, this was not the case for the LDE value, where the increase continued from 30 to 40 repetitions, suggesting that reliability of the LDE may further increase with >40 repetitions. Between-sessions reliability shows a less consistent picture.

4. Discussion

To the best of our knowledge, this is the first study to evaluate within- and between-sessions reliability of measures to characterize lumbar movement patterns, during repeated seated reaching, in a mixed group of participants with and without low back pain. Reliability was calculated for both variability (meanSD and cyclSD), and local dynamic stability (LDE). The results showed moderate within-session reliability for all measures. Between-sessions, the results showed at best moderate reliability.

Several causes can limit the reliability of the measures studied. First, the LDE is known to be affected by fatigue (Asgari et al., 2017; Granata and Gottipati, 2008). However, in our study the mean rating of perceived exertion was similarly low within- and between-sessions (Table 1). Second, before trial 1A, participants were unfamiliar with the experimental procedure. Uncertainty about the burden on the back as a consequence of the task (Lipshitz and Strauss, 1997), might be reflected in their movement patterns (Krüger and Hermsdörfer, 2019). Participants practiced the movement 5 times before trial 1A, but had not yet experienced the total experimental procedure. In line with this, post hoc analysis showed a significant Spearman correlation of pain intensity with meanSD flexion–extension ($r = 0.598$, $p \leq 0.019$) and with LDE ($r = 0.518$, $p \leq 0.048$), only for the first trial in the people with LBP. An effect of uncertainty could then contribute to differences in cyclSD and LDE, between trials 1A and 2, which is also reflected in the lower between-sessions ICC values between these trials, as compared to 1B and 2. Therefore, in multi-session trials, a trial run is recommended. Third, variability in positioning of the participants between trials may have caused variability between measurements. We tried to minimize this error by marking the sitting position on the chair. Contributing to variance between days, the markers were reattached on day 2. Small differences in markers position on the back cannot be ruled out. Fourth, pain affects movement (Hodges and Smeets, 2015), and pain may affect the measures studied (Graham et al., 2014; Asgari et al., 2015). However, pain intensity was comparable between days. Mean absolute pain difference between days was 1.3 (SD 1.3). Fifth, ICC-values of the LDE are negatively influenced by differences in movement speed (Granata and England, 2006). We decided not to use a metronome, to avoid influencing the natural way of moving.

Table 1
Participant characteristics.

36 participants: LBP (n = 15) and Healthy (n = 21)			
Age (year)	35.8 (SD 13.6)		
BMI	24.0 (SD 3.0)		
Gender (M/F)	15/21		
Days between trial 1/2 and trial 3	7.9 (SD 6.4)		
Rating of perceived exertion after trial	1.6 (SD 1.8) trial 1A	1.8 (SD 1.9) trial 1B	1.8 (SD 1.9) trial 2
STBST LBP Participants	1.7 (SD 1.6)		
ODI LBP Participants	17.7 (SD 15.2)		
Pain intensity at test moment LBP Participants	2.3 (SD 1.4) day 1		2.5 (SD 1.8) day 2

BMI Body Mass Index, *STBST* StarT Back Screening Tool, *ODI* Oswestry Disability Index.

Table 2
Test scores, mean differences, p values One-Sample T Test, $\sigma_{differences}$, Limits of agreement and Coefficient of variation at 40 repetitions.

	Mean score trial 1A (mean ± SD)	Mean score trial 1B (mean ± SD)	Mean score trial 2 (mean ± SD)	Trials	Mean difference	One-Sample T Test p value	$\sigma_{difference}$	LoA**	Coefficient of Variation %
MeanSD flexion–extension (degrees)	1.80 (0.7)	1.71 (0.5)	1.77 (0.7)	1A-1B	0.087	0.251	0.45	−0.79, 0.96	16
				1A-2	0.040	0.649	0.51	−0.96, 1.04	17
				1B-2	−0.047	0.568	0.48	−1.00, 0.90	18
MeanSD axial-rotation (degrees)	0.89 (0.3)	0.91 (0.3)	0.90 (0.3)	1A-1B	−0.019	0.585	0.21	−0.43, 0.39	14
				1A-2	−0.001	0.980	0.31	−0.62, 0.61	19
				1B-2	0.019	0.704	0.29	−0.54, 0.58	18
MeanSD lateral-bending (degrees)	0.90 (0.3)	0.89 (0.3)	0.85 (0.2)	1A-1B	0.010	0.826	0.26	−0.49, 0.51	17
				1A-2	0.056	0.145	0.22	−0.38, 0.49	17
				1B-2	0.046	0.296	0.26	−0.46, 0.55	18
CyclSD *(seconds)*	0.15 (0.1)	0.14 (0.1)	0.12 (0.0)	1A-1B	0.016	**0.025***	0.04	−0.07, 0.10	17
				1A-2	0.031	**0.000***	0.05	−0.06, 0.12	22
				1B-2	0.015	0.067	0.05	−0.08, 0.11	19
LDE	4.02 (0.3)	3.95 (0.3)	3.88 (0.3)	1A-1B	0.066	0.143	0.26	−0.45, 0.58	5
				1A-2	0.146	**0.021***	0.36	−0.55, 0.85	6
				1B-2	0.089	0.063	0.28	−0.45, 0.63	5

* Significant difference.
** LoA: Limits of agreement: mean difference ± 1.96 × $\sigma_{difference}$.

Table 3
Intraclass Correlation Coefficients (ICC) within- and between-session for meanSD and local dynamic exponent at 40 repetitions.

	Within-session		95% Confidence Interval		Between-session		95% Confidence Interval	
	Trial	ICC	Lower Bound	Upper Bound	Trial	ICC	Lower Bound	Upper Bound
MeanSD$_{flexion-extension}$	1A-1B	0.71*	0.51	0.84	1A-2	0.73	0.52	0.85
					1B-2	0.69	0.47	0.83
MeanSD$_{axial-rotation}$	1A-1B	0.73*	0.52	0.85	1A-2	0.47	0.16	0.69
					1B-2	0.51	0.21	0.72
MeanSD$_{lateral-bending}$	1A-1B	0.60*	0.34	0.78	1A-2	0.65	0.41	0.81
					1B-2	0.44	0.14	0.67
CyclSD	1A-1B	0.68*	0.44	0.82	1A-2	0.45	0.08	0.69
					1B-2	0.56	0.29	0.75
LDE	1A-1B	0.62*	0.37	0.79	1A-2	0.25	−0.05	0.53
					1B-2	0.54	0.26	0.74

* p value ≤ 0.05.

For variability, the present study showed, consistent, moderate ICC values for meanSD$_{flexion-extension}$ (ICC = 0.69–0.73). This is higher than previously reported ICC value of 0.49 (between-sessions) for standing repeated flexion–extension movements in a group of LBP patients (Graham et al., 2020). This also applies to the ICC value for meanSD$_{lateral-bending}$: 0.44–0.65 in the present study versus 0.29 reported by Graham et al. (2020). The ICC value for meanSD$_{axial-rotation}$ (0.47–0.51) was lower in the present study, compared to 0.75 reported by Graham et al. (2020). In addition to differences in posture (standing vs sitting), these studies used different instrumentation for data collection (inertial measurement units (IMU's) vs a motion capture system). In the present study, the between-sessions ICC values for meanSD$_{flexion-extension}$, exceeded the ICC values for meanSD$_{lateral-bending}$ and meanSD$_{axial-rotation}$. Since the amplitude of lumbar flexion–extension exceeded the amplitude of the lumbar torsion and latero-flexion, this could be attributed to a better signal to noise ratio.

For the LDE, good within-session repeatability in healthy participants (ICC = 0.87 for repeated standing lumbar flexion/extension) has been reported (Dupeyron et al., 2013). In this study, participants performed 100 consecutive repetitions and the first 50 repetitions were compared to the second 50 repetitions. In the present study, the participants stepped off the platform for a minimum pause of 10 min between trials 1A and 1B. Although the procedure used in the earlier study increases ICC values, reliability should be tested over separate test moments, to be generalizable to longitudinal studies. Other differences between these studies are the posture (standing vs sitting), participants (healthy vs a population including patients), and the number of repetitions (50 vs 40). Another study (Graham et al., 2020), reported a between-sessions

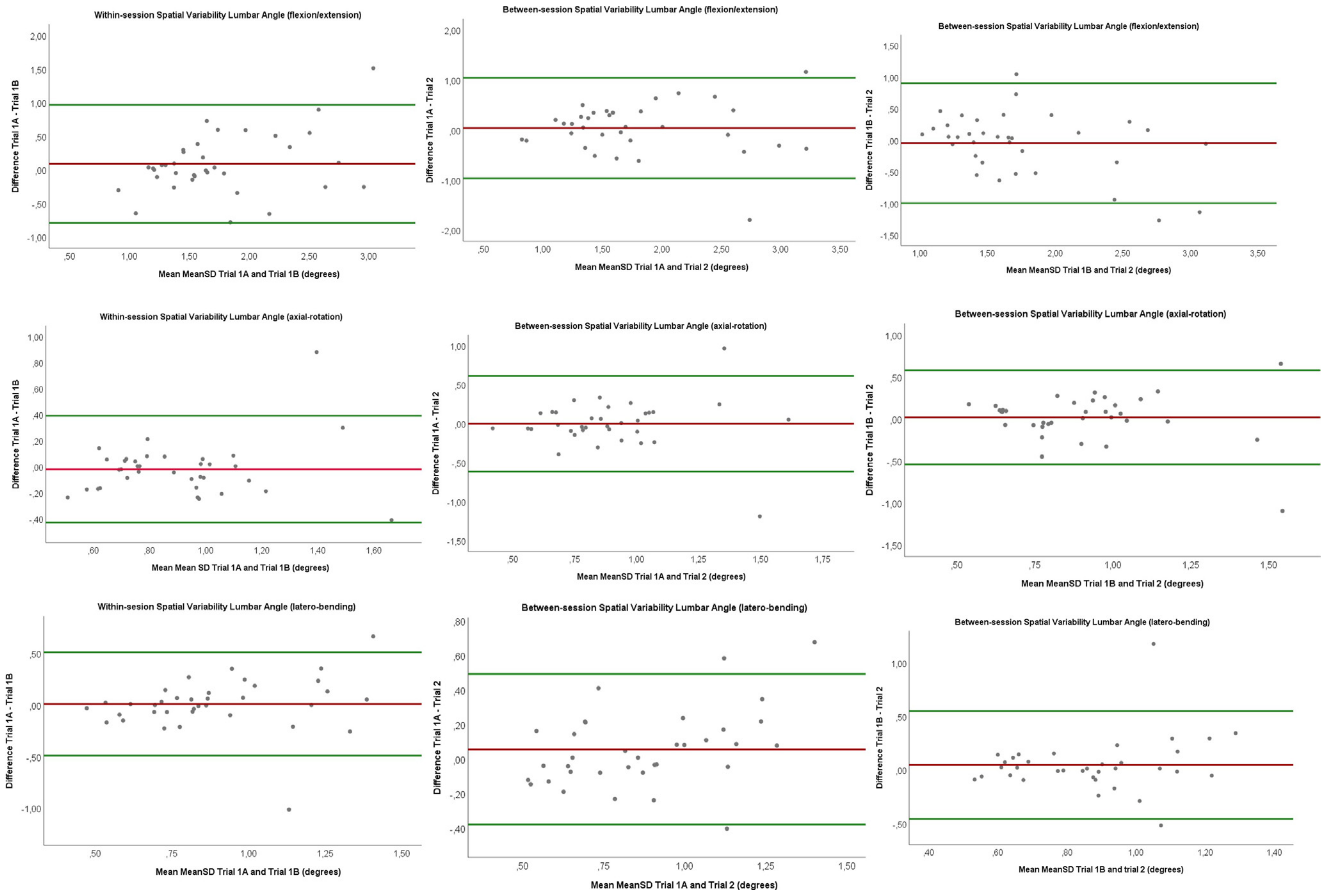

Fig. 3. Bland-Altman plots for within- and between-session reliability of the meanSD (three degrees of freedom).

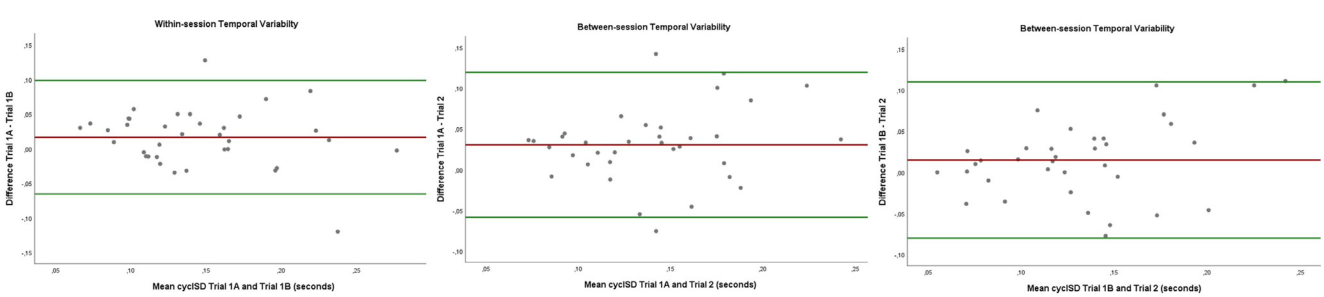

Fig. 4. Bland-Altman plots for within- and between-session reliability of the temporal variability (CyclSD).

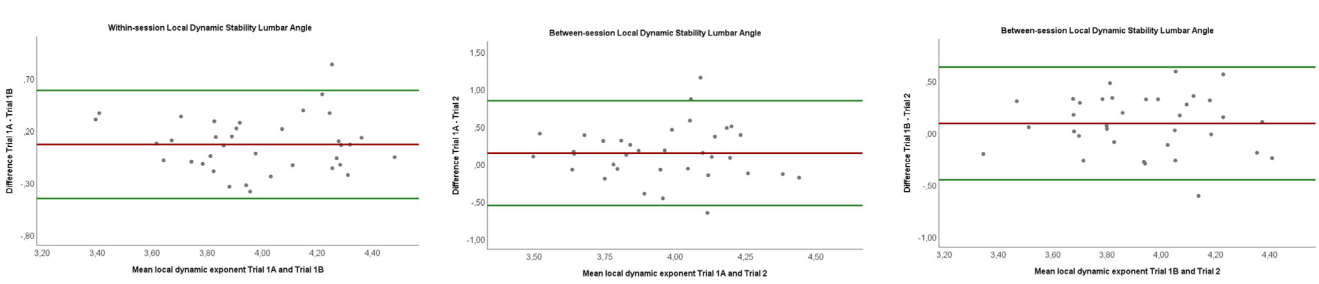

Fig. 5. Bland-Altman plots for within- and between-session reliability of the LDE.

ICC of 0.49 for the LDE in repeated standing lumbar flexion/extension in LBP patients, comparable to the ICC of 0.54 reported here between measurements 1B and 2.

In general, reliability increased with the number of repetitions as expected. This is comparable to the results of Dupeyron et al. (2013). Also in this present study, with the successive addition of

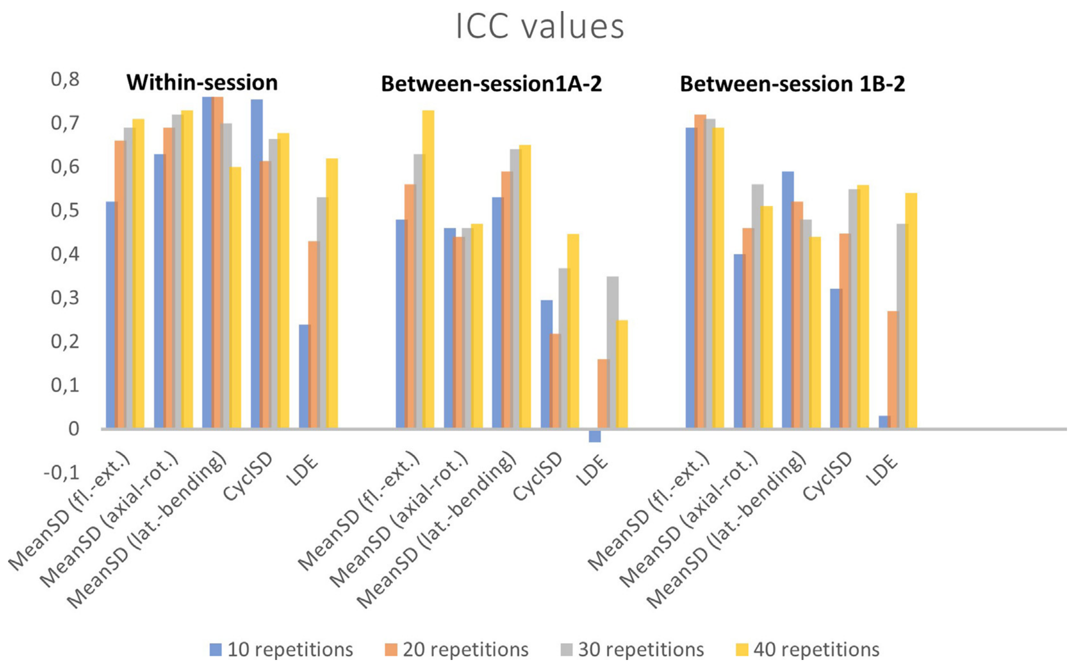

Fig. 6. Number of repetitions of the seated reaching task and corresponding ICC values for meanSD (three degrees of freedom), TVar and LDE, within- and between-session.

10 repetitions in the analysis of the timeseries (starting with 10), the increase of the ICC values seemed to reach a plateau. In our study, 40 repetitions resulted in optimal reliability without fatigue or pain, in a population consisting of people with and without low back pain.

We applied parametric analyses to all variables. With large enough sample sizes (>30), parametric procedures can be used, even when the data are not normally distributed (Pallant, 2007; Elliott and Woodward, 2007). For variables violating the assumption of normality according to the Shapiro-wilk test, non-parametric test were additionally performed. The non-parametric results were comparable or showed higher values than the ICC-values. Based on this we chose the more conservative parametric method.

As reflected in the low coefficient of variation for the LDE, and substantially smaller limits of agreement compared to meanSD and cyclSD, there was low variance between the measurements of the local dynamic stability in this mixed participant group. In relation to statistical power of future experiments on changes in lumbar movement patterns in response to (the threat of) of unpredictable perturbations, the coefficient of variation would be a more relevant indicator than the ICC. Future studies should focus on the relation between trunk movement patterns and potential determinants of changes in movement patterns, such as proprioceptive impairments, and fear of pain, specifically in people with pain.

5. Conclusions

Within-session, the results if this study confirmed the hypothesis. ICC's were moderate and CV values < 20% for meanSD, CyclSD and LDE, indicating that these measures are sufficiently reliable to assess movement patterns of the lumbar spine, in single-session experiments, in a mixed group of people with and without LBP. Between-sessions, the hypothesis was also confirmed, although reliability was lower, it was sufficient to assess movement patterns of the lumbar spine in multi-session experiments. Within-session, except for $\text{meanSD}_{\text{lateral-bending}}$, the number of repetitions had a positive effect on reliability, and a plateau in reliability appears to be reached at 40 repetitions for $\text{meanSD}_{\text{flexion-extension}}$, $\text{meanSD}_{\text{axial-rotation}}$, and cyclSD. Between-sessions, the effect of the number of repetitions on reliability was less consistent.

Funding

This work was supported by the Dutch Organization for Scientific Research: NWO, The Hague, The Netherlands [grant number 0.23 0.12 0.25].

Declaration of Competing Interest

All authors wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgments

The authors are grateful to the Dutch Organization for Scientific Research (NWO, The Hague, The Netherlands) for financial support of this research.

References

Asgari, N., Sanjari, M.A., Esteki, A., 2017. Local dynamic stability of the spine and its coordinated lower joints during repetitive Lifting: Effects of fatigue and chronic low back pain. Human Movement Sci. 54, 339–346.
Asgari, M., Sanjari, M.A., Mokhtarinia, H.R., Sedeh, S.M., Khalaf, K., Parnianpour, M., 2015. The effects of movement speed on kinematic variability and dynamic stability of the trunk in healthy individuals and low back pain patients. Clin. Biomech. 30, 682–688.
Bruijn, S.M., Dieën van, J.H., Meijer, O.G., Beek, P.J., 2009. Statistical precision and sensitivity of measures of dynamic gait stability. J. Neurosci. Methods 178, 327–333.
Bruijn, S.M., Meijer, O.G., Beek, P.J., Dieën van, J.H., 2013. Assessing the stability of human locomotion: a review of current measures. J. Roy. Soc. Interface 10, 20120999. https://doi.org/10.1098/rsif.2012.0999.
De Vet, H.C.W., Terwee, C. B., Mokkink, L. B., Knol, D.L., 2011. Measurement in Medicine, Cambridge University Press. ISBN 978-0-521-11820-0.
Dingwell, J.B., Marin, L.C., 2006. Kinematic variability and local dynamic stability of upper body motions when walking at different speeds. J. Biomech. 39, 444–452.

Dupeyron, A., Rispens, S.M., Demattei, C., Dieën van, J.H., 2013. Precision of estimates of local stability of repetitive trunk movements. Eur. Spine J. 22, 2678–2685.

Elliott, A.C., Woodward, W.A., 2007. Statistical Analysis Quick Reference Guidebook with SPSS Examples. Sage Publications, London.

Fairbanks, J.C., Couper, J., Davies, J.B., O'Brien, J.P., 1980. Oswestry low back pain disability questionnaire. Physiotherapy 66 (8), 271–273.

Graham, R.B., Dupeyron, A., van Dieën, J.H., 2020. Between-day reliability of IMU-derived spine control metrics in patients with low back pain. J. Biomech. 113, 110080. https://doi.org/10.1016/j.jbiomech.2020.11008000.

Graham, R.B., Oikawa, L.Y., Ross, G.B., 2014. Comparing the local dynamic stability of trunk movements between varsity athletes with and without non-specific low back pain. J. Biomech. 47, 1459–1464.

Graham, R.B., Sadler, E.M., Stevenson, J.M., 2012. Local dynamic stability of trunk movements during the repetitive lifting of loads. Human Movement Sci. 31, 592–603.

Granata, K.P., England, S.A., 2006. Stability of dynamic trunk movement. Spine 31, 1–13.

Granata, K.P., Gottipati, P., 2008. Fatigue influences the dynamic stability of the torso. Ergonomics 51, 1258–1271.

Hartvigsen, J., Hancock, M.J., Kongsted, A., Louw, Q., Ferreira, M.L., Genevay, S., Hoy, D., Karppinen, J., Pransky, G., Sieper, J., Smeets, R.J., Underwood, M., 2018. What low back pain is and why we need to pay attention. Lancet 391, 2356–2367.

Hill, J.C., Dunn, K.M., Lewis, M., Mullis, R., Main, C.J., Foster, N.E., Hay, E.M., 2008. A primary care back pain screening tool: Identifying patient subgroups for initial treatment. Arthritis Care Res. 59, 632–641.

Hodges, P.W., Smeets, R.J., 2015. Interaction between pain, movement, and physical activity: Short-term benefits, long-term consequences, and targets for treatment. Clin. J. Pain 31, 97–107.

Kang, H.G., Dingwell, J.B., 2006. Intra-session reliability of local dynamic stability of walking. Gait Posture 24, 386–390.

Koo, T.K., Li, M.Y., 2016. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropractic Med. 15, 155–163.

Krüger, M., Hermsdörfer, J., 2019. Target uncertainty during motor decision-making: The time course of movement variability reveals the effect of different sources of uncertainty on the control of reaching movements. Front. Psychol. 10, 1–13.

Lipshitz, R., Strauss, O., 1997. Coping with uncertainty: A naturalistic decision-making analysis. Organ. Behav. Hum. Decis. Process. 69, 149–163.

Maaswinkel, E., Griffioen, M, Perez, R.S.G.M., Van Dieen, J.H., 2016. *Methods for assessment of trunk stabilization, a systematic review*. J. Electromyogr. Kinesiol. 26, 18–35.

Marques, N.R., Hallal, C.Z., Spinoso, D.H., Morcelli, M.H., Crozara, L.F., Goncalves, M., 2017. Applying different mathematical variability methods to identify older fallers and non-fallers using gait variability data. Aging Clin Exp Res 29, 473–481. https://doi.org/10.1007/s40520-016-0592-8.

Mavor, M.P., Graham, R.B., 2015. Exploring the relationship between local and global dynamic trunk stabilities during repetitive lifting tasks. J. Biomech. 48, 3955–3960.

Mehdizadeh, S., 2019. A robust method to estimate the largest Lyapunov exponent of noisy signals : A revision to the Rosenstein's algorithm. J. Biomech. 85, 84–91. https://doi.org/10.1016/j.jbiomech.2019.01.013.

Pallant J., 2007. SPSS survival manual, a step by step guide to data analysis using SPSS for windows. 3rd ed. Sydney: McGraw Hill, Open University Press, New York, 179-200. ISBN-10: 0335223664.

Owen, N., Bauman, A., Brown, W., 2009. *Too much sitting: a novel and important predictor of chronic disease risk?*. Br. J. Sports Med. 43 (2).

Reynard, F., Terrier, P., 2014. Local dynamic stability of treadmill walking: Intrasession and week-to-week repeatability. J. Biomech. 47, 74–80.

Stergiou, N., Decker, L.M., 2011. Human movement variability, nonlinear dynamics, and pathology: Is there a connection?. Human Movement Sci. 30, 869–888.

Van Dieën, J.H., Reeves, N.P., Kawchuk, G., Dillen van, L.R., Hodges, P.W., 2019a. Analysis of Motor control in patients with low back pain: A key to personalized care?. J. Orthopaedic Sports Phys. Therapy 49, 380–388.

Van Dieën, J.H., Reeves, P.N., Kawchuk, G., Dillen van, L.R., Hodges, P.W., 2019b. Motor Control changes in low back pain: divergence in presentations and mechanisms. J. Orthopaedic Sports Phys. Therapy 49, 370–379.

Van Schooten, K.S., Rispens, S.M., Pijnappels, M., Daffertshofer, A., Dieën van, J.H., 2013. Assessing gait stability: The influence of state space reconstruction on inter- and intra-day reliability of local dynamic stability during over-ground walking. J. Biomech. 46, 137–141.