Manual Therapy 24 (2016) 81–84

Contents lists available at ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/math

Technical and measurement report

Reliability of lumbar movement dysfunction tests for chronic low back pain patients

Christoph Michael Bauer [a, b, *, 1], Martin Heimgartner [a, c, 1], Fabian Marcel Rast [a], Markus Josef Ernst [a], Sarah Oetiker [a], Jan Kool [d]

[a] Institute of Physiotherapy, Department of Health, Zurich University of Applied Sciences, Technikumstrasse 71, 8400 Winterthur, Switzerland
[b] University of Tampere, School of Medicine, Kalevantie 4, FI-33014 University of Tampere, Tampere, Finland
[c] aarReha Schinznach, Fachklinik für Rehabilitation, Rheumatologie und Osteoporose, 5116 Schinznach-Bad, Switzerland
[d] Rehabilitationszentrum Valens, 7317 Valens, Switzerland

ARTICLE INFO

Article history:
Received 8 October 2015
Received in revised form
22 February 2016
Accepted 24 February 2016

Keywords:
Generalizability-theory
Movement disorders
Reproducibility of results
Biomechanical phenomena
Low back pain

ABSTRACT

Assessment of lumbar movement dysfunction commonly comprises trunk range of motion (ROM), movement or control impairment (MCI), and reposition error (RE). Those assessments are typically based on visual observation. Consequently it is not possible to reliably quantify back movements for inter-subject comparisons, or for monitoring changes before and after an intervention. Inertial measurement unit (IMU)-systems could be used to quantify these movement dysfunctions in clinical settings. The aim of this study was to evaluate the reliability of movement dysfunction tests when measured with a novel IMU-system. The reliability of eleven movement dysfunction tests (four ROM, six MCI and one RE tests) were analysed using generalizability-theory and minimal detectable change, measuring 21 chronic low back pain patients in seven trials on two days. Reliability varied across tests and variables. Four ROM and selected MCI tests and variables were identified as reliable. On average, ROM test were more reliable, compared to MCI and RE tests. An attempt should be made to improve the reliability of MCI and RE measures, for example through better standardizations. Subsequently these measures should be studied further for intersubject comparisons and monitoring changes after an intervention.

1. Background and purpose

Low back pain (LBP) is a common disorder with a lifetime prevalence as high as 84% and a high probability of recurrence (Airaksinen et al., 2006; Costa Lda et al., 2009). Contemporary LBP triage systems propose that there is a large group of patients who present with movement dysfunctions (MD), which are a relevant and a provocative factor for ongoing pain (O'Sullivan, 2005; Vibe Fersum et al., 2009). Tests for MD are specifically comprised of 1) range of motion (ROM) (Laird et al., 2014), 2) movement control impairments (MCI) (Sahrmann, 2002; Luomajoki et al., 2007) and 3) tests for proprioception deficits such as reposition error tests (RE) (Rausch Osthoff et al., 2015). These tests typically consist of visual observation (Oesch et al., 2007) and do not quantify MD for diagnostic and outcome evaluation purposes (Seffinger et al., 2004; van Trijffel et al., 2005; May et al., 2006; Stochkendahl et al., 2006; Littlewood and May, 2007).

To overcome these limitations, wireless movement analysis systems using body-worn sensors have recently been developed (e.g. Valedo® from Hocoma AG, ViMove from dorsaVi, or Reablo® from Corehab). These clinical systems comprise of multiple small light weight inertial measurement units (IMU) (Roetenberg et al., 2007). By combining the output of multiple IMU's and post processing algorithms into an IMU-system it is possible to estimate joint angles in a non-invasive way. In a previous study one IMU-system, consisting of four IMUs, was found to be concurrently valid for measures of trunk and hip movement (Bauer et al., 2015).

One prerequisite for tests on MD is high reliability. Four ROM tests and two MCI tests were found to have high reliability, in an asymptomatic population, when measured with an IMU-system (Bauer et al., 2015). However reliability is dependent on the

* Corresponding author. Institute of Physiotherapy, Department of Health, Zurich University of Applied Sciences, Technikumstrasse 71, 8400 Winterthur, Switzerland.
E-mail addresses: christoph.bauer@zhaw.ch (C.M. Bauer), martin.heimgartner@aarreha.ch (M. Heimgartner), fabian.rast@zhaw.ch (F.M. Rast), markus.ernst@zhaw.ch (M.J. Ernst), sarah.schelldorfer@zhaw.ch (S. Oetiker), jan.kool@kliniken-valens.ch (J. Kool).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.math.2016.02.013
1356-689X/

heterogeneity of the sample and is therefore only applicable for a population with a similar heterogeneity (de Vet et al., 2006). Data on reliability of MD tests, when measured with the IMU-system, in patient populations such as CLBP patients, is currently lacking. This study assesses the reliability of ROM, MCI, and RE tests, in a population with CLBP and gives recommendations for reliable measurement protocols.

2. Methods

2.1. Participants

Twenty-three CLBP patients were recruited from a rehabilitation centre. Participants were between 18 and 65 years old and were suffering from CLBP for more than twelve weeks. Exclusion criteria were serious pathologies such as non-healed fractures, anomalies, tumours, specific LBP with neurological signs (muscle weakness, sensation or reflex loss) and acute trauma. Participants had to be able to understand German. The regional ethics committee granted approval. All participants gave their written informed consent.

2.2. Measurement system

Four IMUs (Valedo®) were placed on the right thigh (THI), over the sacrum (S2), and at the level of L1 (L1), and T1 (T1), as described elsewhere (Ernst et al., 2013; Schelldorfer et al., 2015) (Fig. 1). The IMUs were mounted on a plastic frame and attached to the skin with hydrogel tape (KCI Medical GmbH 8153 Rümlang, CH). The IMU's contain a tri-axillar gyroscope, magnetometer, and accelerometer, as well as wireless antenna and signal processing unit. IMU sensor data were transmitted to a recording computer with Valedo® Research software, with a 50 Hz sampling frequency. The raw IMU sensor data was transformed into quaternions according to Madgwick et al. (2010). The angular difference between two IMU's placed above the body segments was calculated and transformed into tilt/twist angles (Crawford et al., 1999). A complete description of the data processing from raw data to tilt/twist angles is documented elsewhere (Bauer et al., 2015). Data processing and statistics were performed using Matlab™.

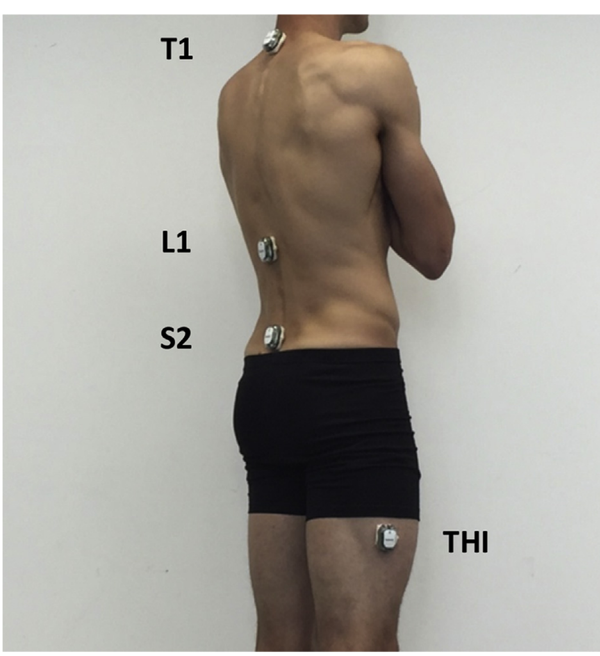

Fig. 1. Experimental setup: IMUs were placed on the right thigh (THI), and level of sacrum (S2), L1 (L1) and T1 (T1).

2.3. Test protocol

Participants attended two identical measurement sessions and performed eleven active movement tests twice within eight days. For retest, the IMU position was marked during the first measurement using a waterproof pen. Test and retest of a participant were conducted by the same examiner. Each session consisted of four ROM, six MCI and one RE tests (Table 2). The order of the tests was randomized between participants but not between sessions. Each test was repeated five times. It took a participant approximately 30 min to perform all tests with five repetitions. The participants did not, to the investigators knowledge, practice the tests between the IMU testing. They were instructed to not alter their routines while they participated in this study. Tests of ROM measure the flexibility of the participant's spine to the end of active range. Tests of MCI evaluate the participant's ability to differentiate movement between two body segments, to stabilize their spine and to move smoothly. These features were analysed by calculating the ratio of the ROM of the respective body segments, by measuring the ROM of the lumbar spine and by the root mean squared jerk (RMSJ). Jerk is defined as the rate of change of angular acceleration and quantifies smoothness of movement (Slaboda et al., 2005). Tests for RE evaluate the participant's proprioceptive deficits within the spine, analysed using absolute error (AE) and constant error (CE) (Rausch Osthoff et al., 2015). Prior to each test the participants received standardized oral instructions by one of the examiners and visual instructions in a video. In case of poor initial performance these instructions were repeated up to three times and the test was demonstrated by one examiner. If the participant was still performing the test incorrectly it was permitted. The participants were instructed to perform the tests at their own preferred speed. Detailed test descriptions and illustrations are provided in supplementary file.

2.4. Statistics

Generalizability theory was used to estimate reliability (Brennan, 2001), with the design $p \times t \times d$ (*participants* $\times$ *trials* $\times$ *days*) based on the linear model

$$X_{ptd} = \mu + \nu_p + \nu_t + \nu_d + \nu_{pt} + \nu_{pd} + \nu_{td} + \nu_{ptd}$$

with μ representing the global mean and ν any one of the seven components.

The index of dependability Φ was calculated as:

$$\Phi = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_t^2}{n_t} + \frac{\sigma_d^2}{n_d} + \frac{\sigma_{pt}^2}{n_t} + \frac{\sigma_{pd}^2}{n_d} + \frac{\sigma_{td}^2}{n_t n_d} + \frac{\sigma_{ptd}^2}{n_t n_d}}$$

with σ being the variance, and n the number of the corresponding component (with n_t and n_d being the number of trials and days). N_t and n_d were equal to one to establish the reliability of a single trial. Φ was interpreted as: <0.25 very low, 0.26–0.49 – low, 0.50–0.69 – moderate, 0.70–0.89 – high, and >0.90 – very high reliability (Carter et al., 2005). Subsequently, Φ coefficients were calculated for alternative measurement strategies, where n_t was varied up to ten trials, and n_d varied across two days, which represent acceptable measurement strategies. Thereby, the number of required trials per day to achieve high reliability was evaluated. High reliability was interpreted as sufficient to compare between different individuals.