Muscular Performance Assessment of Trunk Extensors: A Critical Appraisal of the Literature

Christophe Demoulin* et al.
Department of Motricity Sciences,
Liege University and Liege University Hospital Centre (CHU),
Belgium

1. Introduction

Despite growing research efforts, non-specific low back pain (LBP) remains a major public health burden throughout the industrialized world. Epidemiological data indicate a point prevalence ranging from 19% (Hillman et al., 1996) to 27% (Picavet & Schouten, 2003) and a lifetime prevalence of about 60% (Hillman et al., 1996). Costs to society stem mainly from chronic forms, which account for only 5–10% of cases (Nachemson et al., 2000).

Some literature suggests that muscle dysfunction or increased fatigability might jeopardize the function of the spine and be a risk factor in the development, persistence or recurrence of LBP (Biering-Sorensen, 1984; Parnianpour et al., 1988; Alaranta et al., 1995; Hides et al., 1996). Besides, several studies suggest that patients with chronic low back pain (CLBP) may benefit from an active multidisciplinary approach involving an individually tailored reconditioning program (Bendix et al., 1998; Smeets et al., 2008; Demoulin et al., 2010); some authors even reported benefits of programs based mainly on trunk muscles training (Mannièche et al., 1988; Mooney et al., 1995; Nelson et al., 1995; Carpenter & Nelson, 1999; Mannion et al., 1999b). As a result, tests of trunk muscle performance are essential to get insight in the muscle strength/endurance. Furthermore, accurate evaluation of patients’ deficiencies is essential for the planning of a successful rehabilitation program, for documenting program efficacy and for providing the patients with information on their physical potential and ability to make progress, thereby leading to favourable behavioural changes. Therefore, several reviews have been published targeting performance of trunk muscles (Reimborn & Morrissey, 1988; Newton & Waddell, 1993; Moreau et al., 2001; Malliou et al., 2006). Currently, assessments are performed by means of various methods and no consensus has been reached regarding the optimal test to be used. Most of the time, assessment of trunk extensors has been performed by means of maximum effort tests;

---


1 Department of Motricity Sciences, Liege University and Liege University Hospital Centre (CHU), Belgium
2 Department of Rehabilitation Medicine, School of Caphri, Maastricht University and Maastricht University Medical Centre, Adelante, Centre of Expertise in Rehabilitation and Audiology, The Netherlands
3 Back Unit, Dinant Hospital Centre, Belgium
However, alternatives to maximum effort tests have also been developed. Therefore the aim of the current review is to present a critical appraisal of the literature on this topic.

2. Assessment of trunk extensors by means of maximum effort tests

2.1 Non-dynamometric tests

Trunk extensor performance has been measured with clinical tests for more than 50 years (Hansen, 1964). These tests, which usually assess endurance of trunk extensors, have the main advantages that they don’t require specific equipment, are inexpensive, quick and easy to perform. However, they are not adapted to assess muscle strength and they do not provide a stabilization system to limit hip extensors activation (making them unable to assess spinal muscles specifically). These tests have most often been used in healthy subjects and in patients with CLBP, but they have also been used in other populations (e.g. patients after back surgery (Hakkinen et al., 2003), in schoolchildren (Salminen et al., 1992), etc.).

2.1.1 Static tests

The Sorensen test is by far the most widely used and studied test for assessing trunk extensor muscles (Demoulin et al., 2006b). In this test, the subject lies on an examining table in the prone position with the pelvis aligned with the edge of the table. Calves, thighs, and buttocks are secured and upon command, the subject is asked to maintain the horizontal position as long as possible with the arms folded across the chest (Fig 1a). This test was first described by Hansen in 1964 (Hansen, 1964), but it became known as the “Sorensen test” following a study by Biering-Sorensen in 1984, according to which good isometric endurance might prevent first-time LBP occurrence (Biering-Sorensen, 1984). Although some authors have reported similar findings (Alaranta et al., 1995; Adams et al., 1999; Sjolie & Ljunggren, 2001), such association was not confirmed in other studies (Salminen et al., 1995; Gibbons et al., 1997b; Hamberg-van Reenen et al., 2006).

Since 1984, the Sorensen test has been used in several studies, either in its original or in adapted versions: the differences concerned the arm position, number of straps, criteria for stopping the test, etc. (Demoulin et al., 2006b); the test has also been performed on a roman chair (Fig 1b) in a few studies (Hultman et al., 1993; Mayer et al., 1993; Keller et al., 2001), sometimes with 45 degrees of hip flexion (Champagne et al., 2008). These numerous
methodological variations can affect muscle activity considerably (Mayer et al., 1999; Champagne et al., 2008) and result in considerable discrepancies in study findings. However, concordance was found between some studies regarding the mean holding time in healthy subjects: whereas Latimer et al. measured a holding time of 133s in mixed males and females (Latimer et al., 1999), Mannion et al. reported a holding time reaching 142s and 116s in females and males, respectively (Mannion & Dolan, 1994). Such a gender-related difference was reported in most other studies (Biering-Sorensen, 1984; Mannion et al., 1997a; Kankaanpaa et al., 1998a; McGill et al., 1999; Muller et al., 2010). Differences between genders regarding the weight of the upper body, the degree of lumbar lordosis, the muscles composition (Demoulin et al., 2006b) and the neuromuscular activation patterns (Lariviere et al., 2006) are all hypotheses mentioned.

The Sorensen test has sometimes been considered as a specific tool for evaluating the back muscles (Alaranta et al., 1995). Although spinal muscles are really solicited, most notably the multifidus muscle (Ng et al., 1997; Coorevits et al., 2008; Muller et al., 2010), the test solicits also the other muscles involved in extension of the trunk i.e. the hip extensor muscles (Kankaanpaa et al., 1998a; Plamondon et al., 2002; Plamondon et al., 2004; Champagne et al., 2008; Coorevits et al., 2008; Muller et al., 2010). However muscle fatigue of the hip extensor muscles (reflected by electromyographic parameters) is less correlated to the test holding time than back muscle fatigue (Coorevits et al., 2008).

Although some authors call it a “strength test” (Salminen et al., 1992; Tekin et al., 2009), it rather assesses muscle static endurance (Crowther et al., 2007). Indeed, the elicited contractions were found to be no greater than 40-52% of the maximal voluntary contraction (MVC) (Mannion & Dolan, 1994; Ng et al., 1997; Plamondon et al., 1999; Muller et al., 2010) and the electromyographic (EMG) activity of the spinal erector muscles rarely exceeded 40% of its maximal value (Plamondon et al., 1999; Plamondon et al., 2002).

Although the reproducibility of the Sorensen test has been evaluated in several studies, most of these suffered from methodological weakness (Essendrop et al., 2002; Demoulin et al., 2006b). In general, investigations reported a moderate or high1 intra-session, inter-session and inter-tester reproducibility (Simmonds et al., 1998; Latimer et al., 1999; Demoulin et al., 2008b; Gruther et al., 2009), except in case a Roman chair was used (Mayer et al., 1995; Keller et al., 2001). Although the reproducibility is satisfactory in patients with LBP (Simmonds et al., 1998; Latimer et al., 1999) it might be relevant to repeat the test twice (with a 15-minute rest in between) to avoid a learning effect which has been found in some patients (Demoulin et al., 2008b).

Most studies have reported a good discriminative validity of the Sorensen test reflected by a holding-time being significantly lower in patients with LBP compared to healthy subjects (Biering-Sorensen, 1984; Hultman et al., 1993; Simmonds et al., 1998; Latimer et al., 1999; Ljungquist et al., 1999; Arab et al., 2007; Gruther et al., 2009). The safety of the test has also been investigated. A small number of subjects reported back pain during the test (Demoulin et al., 2008b; Demoulin et al., 2009), sometimes resulting in the interruption of the test (Biering-Sorensen, 1984; Latiikka et al., 1995; Latimer et al., 1999); however, no persistent adverse effects have been reported following the test (Simmonds et al., 1998; Demoulin et

---

1 based on the classification of Wind et al. (J Occup Rehabil, 2005, 15(2):253-272) which will also be used in the rest of the chapter.
al., 2008b) and it could even be applied in elderly people (Champagne et al., 2009). In view of the stress induced on the cardiovascular system, the Sorensen test might better be avoided in patients suffering from cardiovascular disease because of a pressure overload of the cardiovascular system (Suni et al., 1998; Demoulin et al., 2009).

The Ito test, sometimes called “prone isometric chest raise test”, has been described in a couple of studies (Shirado et al., 1995b; Ito et al., 1996; Arab et al., 2007; Durmus et al., 2009; Muller et al., 2010); it consists of lifting the upper body while lying prone with a pad under the abdomen, the arms along the sides, the neck flexed as much as possible and the gluteus maximus muscles contracted for stabilizing the pelvis (Fig. 2a) (Shirado et al., 1995b); this position has to be held as long as possible (Ito et al., 1996). Its discriminative power and high reproducibility were reported in the original study (Ito et al., 1996); furthermore, fatigue of the iliocostalis and the multifidi has clearly been linked to the holding time (Muller et al., 2010). Although this test is attractive because it is easy to perform and because it seems to induce less spine loading and limit the risk of lumbar hyperlordosis as compared to the Sorensen test (Ito et al., 1996), no study really confirmed this assumption. Furthermore, a study suggested that this test was less comfortable and more difficult to standardize (with regard to the extent of the upper body lift) than the Sorensen test (Demoulin et al., 2008b). These differences might explain the controversial correlations found when comparing holding times of both tests (Demoulin et al., 2008b; Muller et al., 2010).

The prone double straight-leg raise test has been described for evaluating the isometric endurance of the lower spinal extensor muscles (McIntosh et al., 1998; Moreau et al., 2001). In this test, the subject lies prone with hips extended and the hands underneath the forehead (Fig. 2b). The subject is asked to raise both legs until knee clearance as long as possible. According to Arab et al., this test is as reproducible as the other static endurance tests and has good sensitivity, specificity and predictive values in LBP (Arab et al., 2007). However, information about its validity, safety and responsiveness is lacking.

Fig. 2. a) Ito test, b) Prone double straight-leg raise test

2.1.2 Dynamic tests (“arch-up tests”)

The “arch-up tests”, sometimes considered as dynamic variants of the Sorensen test, are usually used to assess dynamic endurance of trunk extensors. These tests, performed with the subject prone with the torso cantilevered over the edge of a table, consist in flexing the trunk to a specific position (e.g. 30° trunk flexion), then returning to the initial position as
many times as possible at a determined rate of arch-ups per minute (Fig. 3) (Alaranta et al., 1994; Gronblad et al., 1997; Moreland et al., 1997; Udermann et al., 2003). Whereas the static tests have been widely studied, the dynamic tests have received less attention and have been performed in various ways regarding the support (examination table, roman chair, etc.), the range of motion, the rate per minute, etc. As the original Sorensen test, the dynamic tests are not specifically testing the back muscles (Konrad et al., 2001). Although moderate reliability is suggested (Alaranta et al., 1994; Moreland et al., 1997), little is known about the other clinimetric properties of such tests. Furthermore, a recent study, which compared the static Sorensen test with its dynamic variant, revealed that the latter was less comfortable and more difficult to standardize (Demoulin et al., 2008b). In a few studies, the subjects were asked to perform as many repetitions as possible in 30 seconds (Viljanen et al., 1991; Kujala et al., 1996).

2.2 Dynamometric tests

Today, various dynamometric testing machines have been developed to assess trunk muscle performance: these tests allow more complete, precise and specific assessments than the non-dynamometric tests. These measurement systems, also designed to train muscles, differ in terms of contraction mode (static, isotonic, isokinetic), subject position (standing, sitting, lying prone) etc., and generally enable the assessment of several muscular qualities.

2.2.1 Muscle strength tests

MVC tests of trunk extensor muscles have been used in several studies to assess maximal strength in healthy subjects and patients with LBP but also in other populations (e.g. patients following back surgery (Hakkinen et al., 2003), elderly subjects (Rantanen et al., 1997), etc.). Patients with CLBP had reduced values compared to healthy controls in most studies (Reid et al., 1991; Hultman et al., 1993; Kankaanpaa et al., 1998b; Handa et al., 2000; Bayramoglu et al., 2001; Kramer et al., 2005; Gruther et al., 2009), but not all studies (Shirado et al., 1992; Cassisi et al., 1993; Takemasa et al., 1995; da Silva et al., 2005). Several methods (see below) have been used for testing maximal strength.
2.2.1.1 Static strength test

Usually, after a muscular warming-up and sometimes a familiarization period, the subject is instructed to build up the force with increasing intensity. In most studies, about three MVC are measured at short periods intervals; sometimes additional trials are permitted and the best result of the contractions is selected (Demoulin et al., 2006a; Schenk et al., 2006).

Trunk extensors strength can be assessed by means of a hand-held dynamometer that is held by the investigator in the interscapular area; the subject, lying prone, has to perform a maximal static effort against it (Fig. 4a). This test which has been confidentially described (Moreland et al., 1997; Swezey et al., 2000; Durmus et al., 2009) appears to be difficult to perform in a standardized manner (Moreland et al., 1997; Swezey et al., 2000) and has a low reproducibility (Moreland et al., 1997).

MVC of trunk extensors has also been assessed by means of a strain gauge (Fig. 4b) attached to a wall and connected to a strap around the shoulders; a pelvic fixation is provided so that the rotation axis is set at the hip joint level. The subject, in standing position, has to perform an isometric backward extension (“pulling test”) (Biering-Sorensen, 1984; Nicolaisen & Jorgensen, 1985; Kumar et al., 1995; Kujala et al., 1996). In some studies, a more sophisticated apparatus (e.g. with a frame) has been developed (da Silva et al., 2005). Tests in sitting (Kumar et al., 1995) or in prone positions (Plamondon et al., 2004; da Silva et al., 2005) have also been described. Reliability of the pulling tests seems high to moderate (Jorgensen, 1997; Lariviere et al., 2001); however, little is known about the other clinimetric qualities.

Fig. 4. a) Hand-held dynamometer, b) “Pulling test” in standing position, c) “Pulling test” in prone position

Specialized and commercialized equipments have also been developed to assess and train trunk muscles. The subject is seated in the equipment and a control of the pelvis is provided by means of a stabilization system designed to limit the activation of hip extensors (Graves et al., 1994; San Juan et al., 2005; Smith et al., 2008); however relevance of such stabilization systems which differ from one device to another remains controversial (Udermann et al., 1999; Walsworth, 2004).
Most studies concerned the MedX™ (MedX Corp. Ocala, FL, USA) which is a dynamometer developed to assess and train spinal muscles (Graves et al., 1990) (Fig. 5). MedX™ assessment consists of measuring the extensor isometric MVC at 7 angles of trunk flexion within the patient’s range of motion (i.e. 0-12-24-36-48-60-72°) (Graves et al., 1990). This device is unique in the fact that it uses a gravity correction system (Pollock et al., 1991; Graves et al., 1994). Literature suggests a moderate to high reproducibility of peak torque values in healthy individuals (Graves et al., 1990) and patients with CLBP (Robinson et al., 1992).

Other companies (David®, Tergumed®, Schnell®, DBC®) propose a complete set of four individual units for training (Taimela & Harkapaa, 1996; Daniels & Denner, 1999; Mannion et al., 1999b; Giumza et al., 2006) and assessing the trunk extensor, flexor, rotator and lateral-flexor muscles, respectively (Demoulin et al., 2006a; Roussel et al., 2008). The extension device (Fig. 5) differs between the various systems of the companies regarding the hip stabilization system, position of the thighs, legs and feet, etc. Although these protocol differences might concur meaningful inter-system comparison, significant correlations were observed between the MVCs measured by the David®, Tergumed® and Schnell® systems as well when considering the absolute values ($r \geq 0.8$) as when considering the relative values expressed in percentage of specific normative data ($r \geq 0.69$) (Demoulin et al., 2008a).

Although spinal muscles seem to be well activated ($\geq 80\%$ maximal EMG activity) during an isometric extension MVC on such dynamometers (Denner, 1997; Vanderthommen et al., 2007), a significant activation of hip extensor muscles has also been observed (about $50\%$ of maximal EMG activity) (Vanderthommen et al., 2007). Several authors reported a high intersession (Elfving et al., 1999; Demoulin et al., 2006a) and inter-tester (Demoulin et al., 2006a) reproducibility of MVC measurements in healthy subjects and in patients with CLBP (Elfving et al., 2003; Roussel et al., 2008); however, the inter-site reproducibility (in healthy subjects) revealed small but significant differences in measurements between identical devices (Demoulin et al., 2006a). The cardiovascular stress of such maximal isometric effort seems to be limited in healthy middle-aged individuals (maximal systolic and diastolic blood pressure monitored at the end of the MVC test: 165 and 105 mmHg, respectively) (Demoulin et al., 2009); however these results need to be confirmed with instantaneous blood pressure measurement.
A positive relationship between lifting and LBP has been reported (Cole & Grimshaw, 2003); as a result some functional assessments (lifting test) have been developed to measure the strength of the functional chain (upper limbs-trunk-lower limbs) during static lifting tasks (Newton et al., 1993; Mannion et al., 1997a; da Silva et al., 2005; Ropponen, 2006). While standing and bending forward, the subject is asked to pull upward a handlebar which is fixed by a chain to a floor-mounted load cell. Methods of testing described in the literature differ regarding materials, knee flexion, the bar height, etc. Though it is a lifting task, the real functional aspect of such test remains questionable because it involves no movement; the safety of such lifting maximal isometric task remains also controversial (Hansson et al., 1984). Limited evidence is available about the clinimetric qualities of such tests.

2.2.1.2 Isokinetic test

Isokinetic dynamometry has been one of the most widely used approaches to train and measure strength of trunk muscles (Newton et al., 1993) for more than 30 years (Hasue et al., 1980). Such dynamometers can measure trunk flexion and extension strength (allowing to calculate agonist/antagonist ratios (Newton et al., 1993)(Fig. 6a), at various angular speeds and contraction modes (concentric most often but also eccentric (Shirado et al., 1992) and isometric (Bayramoglu et al., 2001; McGregor et al., 2004; Gruther et al., 2009)). Another advantage of isokinetic dynamometry is that it provides a variable resistance accommodating to a painful arc during the movement. Test-retest reliability of isokinetic measurements appears high in healthy subjects in most studies (Delitto et al., 1991; Newton et al., 1993; Keller et al., 2001; Karatas et al., 2002). In patients with LBP, an increase in performance between test and retest, interpreted as “learning effect”, has often been reported (Grabiner et al., 1990; Newton et al., 1993; Keller et al., 2001; Gruther et al., 2009). Inter-site reliability, tested in healthy volunteers, also seems to be high (Byl & Sadowski, 1993).

However, use of isokinetic dynamometry to perform trunk muscle assessment suffers from several limitations: although some authors tried to propose a standard method of testing (Dvir & Keating, 2001), no consensus has been established yet regarding the optimal parameters for testing i.e. movement speed (which can affect testing accuracy (Keller et al., 2001)), range of motion, number of repetitions (Genty & Schmidt, 2001), etc. Differences between the existing isokinetic trunk testing machines in terms of subject position (sitting vs standing)(Morini et al., 2008), ways to reduce the artefacts, stabilization system, gravity correction system (Hupli et al., 1997; Findley et al., 2000) limit meaningful inter-system comparison (Hupli et al., 1997). Besides, the stabilization systems might be inefficient to avoid involvement of hip muscles, especially in the standing position (Morini et al., 2008). Finally, according to some authors (Ayers & Pollock, 1999), the validity of the isokinetic tests of trunk extensors remain controversial due to the impact forces at the end of the movements which can induce artefacts (overshoot). Furthermore, these tests could induce vagal disturbances (Genty & Schmidt, 2001) and pain during testing (Shirado et al., 1995a; Genty & Schmidt, 2001).

Isokinetic dynamometer has also been used to measure the strength of the functional chain (liftask) (Newton et al., 1993; Latikka et al., 1995; Gibbons et al., 1997b; Ropponen, 2006). As for the static lifting tests, various methods of testing have been described in the literature; besides, though it is a lifting task, the real functional aspect of such test remains
questionable because it involves a movement in a constant speed. Limited evidence is available about the clinimetric qualities of such tests except for reliability which is high in LBP patients and healthy subjects (Newton et al., 1993; Latikka et al., 1995). The high correlations found between the isokinetic flexion-extension and lifting tests suggest that performing both tests is not necessary in clinical practice (Newton et al., 1993).

2.2.1.3 Isoinertial measurements

The Isostation B-200® (Fig. 6b) has been used in a huge number of studies to assess trunk muscle performance but is less used nowadays. In addition to mobility and isometric MVC measurements, this triaxial lumbar dynamometer allows for isoinertial tests (i.e. use of a constant load throughout the range of motion) (Parnianpour et al., 1989b; Gomez et al., 1991; Balague et al., 2010). For the isoinertial flexion-repetition test, the resistance (free weights) is set at a determined percentage of the MVC of flexion (e.g. 25% or 50% (Hutten & Hermens, 1997)) for the sagittal axis and the subject is asked to bend and then return backward as fast as possible (maximum effort) about five times while functional indices (maximal or average velocity, power index and work index) can be simultaneously assessed (Gomez et al., 1991; Rytokoski et al., 1994). This assessment appears to be safe (Newton & Waddell, 1993) and reliable (Rytokoski et al., 1994) as well in healthy persons (Parnianpour et al., 1989a) as in patients with CLBP (Szpalski et al., 1992; Hutten & Hermens, 1997) except for mobility assessments. Unfortunately, axis of rotation of the device is behind the estimated axis for lumbar spine for flexion and extension (Dillard et al., 1991). Furthermore, the device might be inefficient to fully stabilize the pelvis and its relevance to improve functional physical capacity remains controversial (Sachs et al., 1994).

Fig. 6. a) Isokinetic dynamometer, b) Isostation B-200®
2.2.2 Endurance tests

2.2.2.1 Static endurance

Muscle static endurance can be assessed with several dynamometers by measuring the time during which a subject is able to maintain a specific torque level corresponding at a preset relative percentage (often 40-60%) of the MVC previously determined (Jorgensen, 1997; Kankaanpaa et al., 1998b; Udermann et al., 2003; Demoulin et al., 2009). A visual feedback system, displaying the torque in real time, is generally positioned in front of the subject in order to keep a constant torque. This test performed in standing position, used for more than 25 years (Nicolaisen & Jorgensen, 1985), is sometimes considered to be more appropriate than the Sorensen test because it is less sensitive to heterogeneous physiques (Jorgensen, 1997; Kankaanpaa et al., 1998a; da Silva et al., 2005). Demoulin et al. compared a seated endurance test performed on a specific dynamometer (David®) (Fig. 7a) to the Sorensen test in healthy subjects; they reported limited pain in the back during performance of both tests and similar subjective level of exertion and cardiovascular stress (Demoulin et al., 2009). As for the MVC test performed on this device, this seated endurance test induces hip extensors activation in spite of the hip stabilization system (Kankaanpaa et al., 1998b); unfortunately, this endurance test has a low test-retest reliability as well in healthy subjects as in patients with CLBP (Demoulin, 2008). Static endurance of trunk extensors have also been measured while the subject performs a lifting test (Mannion et al., 1997a; da Silva et al., 2005); however, such tests produce less fatigue in the back muscles than the Sorensen or the pulling tests (da Silva et al., 2005).

2.2.2.2 Dynamic endurance

Muscle dynamic endurance can be assessed with dynamometers by measuring the maximal number of repetitions performed with a specific load, with a preset speed and range of motion. The literature reports only few studies using such tests: on the David® device (Fig. 7b), the load used corresponded to \([0.4 \times \text{height (meter)}} \times [0.6 \times \text{Weight (kg)}] \times 0.82\) (Kankaanpaa et al., 1997). This test seems to be less reproducible and well tolerated than the MVC strength and static endurance tests (Demoulin, 2008). Similar tests have been described with the Isostation B-200® (Morlock et al., 1997) and the MedX™ (Udermann et al., 2003).

Fig. 7. a) Static endurance test, b) Dynamic endurance test
2.2.3 Muscle fatigue tests

Muscle fatigue can be defined as “an exercise-induced reduction in the ability of muscle to produce force or power whether or not the task can be sustained” (Enoka & Duchateau, 2008). Fatigue can be calculated by comparing the maximal strength (MVCs) prior and after an exhaustion task; in the study of Al-Obaidi et al., the task consisted in performing as many extension movements as possible against a predefined individual resistance (corresponding at 50% of the pre-MVC) (Al-Obaidi et al., 2003). Plamondon et al. submitted healthy students to an intermittent prone back extension exercise (100 dynamic repetitions) and reported fatigue of trunk extensors according to the decrease of MVC values (14-20%) measured with a strain gauge in a prone position (Plamondon et al., 2004). Corin et al. compared several ways to test muscle fatigue (Corin et al., 2005) but according to our knowledge, no study has really investigated the clinimetric properties of such assessments. The isokinetic dynamometers enable to assess fatigue resistance of trunk extensors by requiring more than 15 repetitions at maximal intensity; the torque decrease (fatigue index) throughout the test is generally considered as a good indicator of fatigue resistance (Cale-Benzoor et al., 1992; Genty & Schmidt, 2001; McGregor et al., 2004). The high cardiovascular stress induced by such tests, which can be an important factor-limiting performance (Rantanen et al., 1995), might explain why they have been poorly investigated; furthermore, dizziness has been reported after such exercise (Peel & Alland, 1990) and a huge increase in heart rate (HR), which could reach 90% of maximal theoretical HR at the end of 20 repetitions, was reported (Rantanen et al., 1995). Therefore, caution is needed when testing patients with suspected heart problems (Rantanen et al., 1995).

Nowadays, for fatigue assessment, the surface electromyography (S-EMG) technique is often used and coupled to the endurance tests previously described, which are most of the time limited in time; thus S-EMG is used as an alternative to maximum effort tests to assess trunk muscle performance (see below).

2.3 Interpretation of results

Maximum effort tests have generally pointed out decreased trunk muscle performance in patients with CLBP. Most authors having observed such changes suggested that they could result from physical deconditioning and the associated alterations in the size (decrease in cross-sectional surface area of spinal muscles), density (fatty infiltration) and structure (fibers size reduction) of the trunk muscles (Hultman et al., 1993; Gibbons et al., 1997b; Raty et al., 1999; Danneels et al., 2000; Barker et al., 2004; Demoulin et al., 2007). However, several more recent papers consider that there is minimal research evidence that patients with CLBP really suffer from disuse, physical deconditioning (Smeets & Wittink, 2007; Verbunt et al., 2010) and morphologic alterations (Crossman et al., 2004; Smeets & Wittink, 2007; Verbunt et al., 2010).

The decrease in performance found in patients could partly result from of a lack of validity of such assessments which require maximal collaboration of subjects to produce a maximal effort in terms of intensity or duration (Newton & Waddell, 1993). Therefore, results can be influenced by several individual confounding factors such as motivation, pain tolerance, competitiveness (Mannion & Dolan, 1994); furthermore pain on exertion, anticipation or fear of pain and reflex inhibition of motor activation can be additional factors resulting in
inability or unwillingness to produce a truly maximal effort in patients with LBP (Menard et al., 1994; Vlaeyen et al., 1995; Crombez et al., 1996; Keller et al., 1999; Rashiq et al., 2003; Rainville et al., 2004; Al-Obaidi et al., 2005; Ropponen et al., 2005; Verbunt et al., 2005; Thomas et al., 2008; Huijnen et al., 2010). These individual factors might explain the absence or low correlations found in some studies between morphologic variables and performance (Parkkola et al., 1993; Gibbons et al., 1997a). They might also explain the significant learning effect observed in some patients, reflected by performance higher at the second trial than at the first one (Grabiner et al., 1990; Newton & Waddell, 1993; Lariviere et al., 2003b; Gruther et al., 2009). Such learning effect might explain partly the increase in trunk extensor performance sometimes observed after only a few training sessions (Mannion et al., 2001; Demoulin et al., 2010). Therefore, such increase in performance should always be interpreted with caution.

Therefore, although several studies reported no or low correlations between pain or disability and trunk extensor performance (Newton et al., 1993; Gronblad et al., 1997; Bayramoglu et al., 2001; da Silva et al., 2005), these maximum effort tests could also be considered as psychophysical test, reflecting in some cases more the fears and pain tolerance than the muscle function. Consequently, the relevance of using such tests in very painful patients is doubtful. Besides, a period of familiarization with the test appears absolutely necessary in order to eliminate the learning effect and the risk to underestimate real performance.

The technique of twitch interpolation seems a research method able to identify the role of non-physiological factors during strength testing (Verbunt et al., 2003). It is based on the registration of a twitch contraction elicited by a supramaximal electrical stimulus delivered to the muscle or nerve during a MVC. The force increment in response to this stimulus reflects the muscle force reserve or the difference between the maximum force that can be generated by the muscle and the maximum voluntary contraction force, in which nonphysiological factors play a role (Verbunt et al., 2003). This technique was used to compare healthy subjects with patients with LBP regarding knee extensor inhibition in a few studies (Suter & Lindsay, 2001; Verbunt et al., 2005); a lower central activation ratio was reported in patients experiencing increased psychological distress and with higher pain intensity (Verbunt et al., 2005).

3. Alternative to maximum effort tests to assess trunk muscle performance

A few studies examined whether trunk extensor strength could be predicted by anthropometric variables (Mannion et al., 1999a; Wang et al., 2005); indeed, such a prediction is of particular interest in patients who cannot perform maximal tests in order to determine appropriate loads for rehabilitation training (Mannion et al., 1999a). If these variables seem to influence muscle performance, their ability to predict accurately muscle capacity remains limited (Lariviere et al., 2003b).

Taimela et al. developed a submaximal dynamic back extension endurance test utilising subjective perception of low back fatigue (Taimela et al., 1998). They reported that the perceived fatigue (assessed by means of a Borg scale every 15 seconds) increased faster in patients with LBP disorders than in healthy subjects and suggested that this test might be a low-risk, low-cost evaluation method for assessing LBP patient when combined with other
clinical data (Taimela et al., 1998). However, according to our knowledge, no other study has used this test.

Mannion et al. conducted a very interesting study to determine whether the twitch superimposition technique could be used to predict maximum force (isometric lifting test) of the spinal muscles from submaximal efforts (Mannion et al., 1997b). Although they reported an excellent curvilinear relationship between twitch force and submaximal force being sustained, they observed that the predicted MVC (extrapolated from the relationship) underestimated the true strength by about 18%. Such difference might result partly from the difficulty in stimulating the spinal muscle mass as a whole. The authors concluded that another testing apparatus and/or subject’s posture might result in a more accurate prediction of maximal force (Mannion et al., 1997b). However, no other studies have used the twitch superimposition technique to predict back muscle maximal force since then.

Surface electromyography (S-EMG) technique is sometimes considered as the best tool to assess objectively trunk extensors muscle function because it enables to investigate and compare simultaneously and specifically several back muscles. Furthermore, this technique can be used during a submaximal and time-limited effort in order to limit the influence of individual factors (motivation, fears, etc.). Therefore S-EMG coupled to the endurance field (Sorensen, etc.) and dynamometric (static or dynamic) tests previously described have been frequently used in the literature (Mannion et al., 1997a; Elfving et al., 2000; Koumantakis et al., 2001, Ng et al., 1997; Kankaanpaa et al., 2005; Demoulin et al., 2007). Some devices such as the Back Analysis System (NeuroMuscular Research Center, Boston University, Boston, USA) were even developed to standardize assessments of back muscle dysfunction (i.e. repeated isometric extensions at a given percentage of the MVC associated to S-EMG monitoring) (De Luca, 1993; Roy et al., 1995). The EMG power spectrum has been widely used to calculate the median frequency (MF), mean power frequency (MPF), as well as their rates of decline during prolonged exercise in order to reflect muscle fatigue (Vollestad, 1997). Several studies observed that EMG fatigue parameters recorded after a prespecified period (often 45-60 seconds) of a fatiguing task were significantly correlated to the parameters monitored at the end of the endurance test (van Dieen et al., 1998; Suter & Lindsay, 2001) as well as to the maximal holding time (Kankaanpaa et al., 1997; Mannion et al., 1997a; van Dieen et al., 1998; Dederding et al., 1999). Furthermore, EMG fatigue parameters could be a better predictor of low back disorder than the maximal holding time (Mannion et al., 1997a).

Though submaximal tests coupled to S-EMG have become very popular, the validity of the EMG submaximal endurance tests performed at a given percentage of the MVC can be questioned. Indeed, the intensity of effort during such tests depends on the factors (motivation, pain, fears, etc.) influencing the MVC test previously performed; the absence of difference in EMG parameters between healthy and patients with CLBP and the smaller decrease in power frequency (reflecting lower fatigue) found in the latter group in some studies could be explained by the underestimation of the patients MVC resulting in a lower load level (Elfving et al., 2003; Lariviere et al., 2003a; Kramer et al., 2005). In order to avoid such a bias and to limit the influence of the anthropometric variables, Lariviere et al. recently proposed a promising assessment based on S-EMG monitoring during intermittent submaximal static contractions (6,5 seconds contraction / 1,5 second rest) performed in a non-commercial trunk dynamometer at a specific intensity (90 N.m) during 5 to 10 minutes...
Their results based on healthy subjects suggest that the EMG indices used in the study could predict absolute endurance as well as strength with the use of a single intermittent and time-limited endurance test (Lariviere et al., 2008b).

Although S-EMG technique appears attractive, it presents some drawbacks. Indeed, EMG results are influenced by many factors including the type, size, and location of the electrodes, the impedance of the source and amplifier, the location of the motor points, the type of contraction, the temperature of the muscle and skin, the force produced by the contraction, the fiber composition, the blood flow and the fat layer thickness (De Luca, 1993). Whereas intra-session reproducibility of EMG parameters seems generally satisfactory (Ng & Richardson, 1996), inter-session and inter-operator reproducibility remains controversial (Peach et al., 1998; Elfving et al., 1999; Danneels et al., 2001). Furthermore, S-EMG might not reliably isolate the activity of the different back muscles (Stokes et al., 2003) and the interpretation of EMG measurements at an individual level remains impossible at the moment because of the considerable inter-individual variability (Elfving et al., 2000; Arnall et al., 2002), thereby limiting its diagnostic usefulness (Pullman et al., 2000; Lariviere et al., 2002). Finally, the absence of standardized EMG protocols prevents from performing several comparative studies.

4. Conclusions
As shown in this review, several methods have been used to assess trunk extensor muscle performance. Unfortunately there is not yet a consensus regarding the optimal test to be used and the present literature review does not enable such a test to be determined. Further studies about the clinimetric properties of the maximal effort tests as well as comparison studies between the various existing tests and tools are needed. Anyway, when using such tests, several methodological cautions are necessary in clinical practice (e.g. a familiarization period to the device and to the test, several trials authorized, etc.) in order to avoid a learning effect; furthermore, results interpretation should always be careful, especially in painful or fearful subjects considering the risk of underestimating the true muscle performance. Additional effort to develop a submaximal test remains essential. Although the S-EMG technique appears to be a key investigation tool for research because individual factors do not influence the outcomes, further investigations are necessary to make the measurement interpretation possible at an individual level.

5. Acknowledgements
The authors thank Prof M. Szpalsky and F. Balagué for providing some pictures as well as A. Depaifve and S. Wolfs for their help, cooperation and valuable assistance.

6. References


