Outcome Measures of Functioning and Physical Activity in Patients with Low Back Pain

Exemplified in Patients Who Undergo Lumbar Fusion Surgery

Max Jakobsson

Department of Orthopaedics,
Institute of Clinical Sciences,
Sahlgrenska Academy,
University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2019

Not everything that can be counted counts,
and not everything that counts can be counted.

WILLIAM BRUCE CAMERON

Outcome Measures of Functioning and Physical Activity in Patients with Low Back Pain

Exemplified in Patients Who Undergo Lumbar Fusion Surgery

Max Jakobsson

Department of Orthopaedics, Institute of Clinical Sciences,
Sahlgrenska Academy, University of Gothenburg,
Gothenburg, Sweden, 2019

ABSTRACT

INTRODUCTION. Chronic low back pain (LBP) can negatively affect health in terms of disability and decreased levels of functioning and physical activity. Chronic LBP due to degenerative disc disease (DDD) is a subgroup of LBP for which lumbar fusion surgery (LFS) is a treatment option. LFS is usually evaluated with patient-reported outcome measures (PROMs) of disability, but physical capacity tasks measuring functioning and accelerometers measuring physical activity can complement the use of PROMs to better understand patients' health.

AIM. To investigate aspects of the measurement of functioning and physical activity in patients with LBP.

METHODS. In Study I, articles on physical capacity tasks for patients with LBP were systematically identified and the level of evidence for the reliability, validity, and responsiveness of the tasks was determined. Studies II–IV included patients with chronic LBP due to DDD scheduled for LFS. In Study II, the responsiveness and minimal important change of four physical capacity tasks were investigated with hypothesis testing and the optimal cutoff point method. In Study III, patients' preoperative level of physical activity was studied with accelerometers. Associations with potential barriers to physical activity were investigated with regression analysis. In Study IV, preoperative predictors of the patients' levels of physical activity and disability six months after surgery were investigated with regression analysis.

RESULTS. Five-repetition sit-to-stand, five-minute walk, 50-foot walk, progressive isoinertial lifting evaluation, and timed up-and-go demonstrated the best evidence for reliability and validity for patients with chronic LBP (Study I). Of these, five-repetition sit-to-stand also showed adequate responsiveness. One-minute stair climbing demonstrated adequate results for both reliability and responsiveness. In Studies II–IV, 118 patients with chronic LBP due to DDD were included. Fifty-foot walk, timed up-and-go, and one-minute stair climbing demonstrated adequate responsiveness while 5-minute walk did not (Study II). Ninety-eight patients did not fulfill the WHO recommendations on physical activity, of whom 32 did not accumulate a single min-

ute of the required 150 minutes per week of physical activity. Moreover, high levels of fear of movement and disability were associated with a low preoperative level of physical activity (Study III). A low preoperative level of physical activity and a high preoperative level of self-efficacy for exercise were predictors of a larger increase in the postoperative physical activity. A high preoperative level of disability and low preoperative levels of pain catastrophizing and self-efficacy for exercise were predictors of a more favorable outcome for disability (Study IV).

CONCLUSIONS. Fifty-foot walk and timed up-and-go showed adequate results for reliability, validity, and responsiveness and are recommended for assessment of functioning in patients with chronic LBP due to DDD undergoing LFS. Future pre- and postoperative interventions targeting fear of movement and disability might increase the level of physical activity in physically-inactive patients. The prediction model of physical activity could possibly be used in clinical practice to predict which patients are in need of extra pre- and postoperative interventions to increase their level of physical activity.

SAMMANFATTNING PÅ SVENSKA

BAKGRUND. Långvarig ländryggssmärta är vanligt förekommande och kan innebära försämrad hälsa i form av nedsatt fysisk funktion och lägre fysisk aktivitetsnivå. Endast en bråkdel av alla med långvarig ländryggssmärta genomgår kirurgi men antalet operationer ökar årligen. Segmentell rörelsesmärta (SRS) är en typ av ländryggssmärta där steloperation av ländryggen är ett behandlingsalternativ. Resultatet av steloperation utvärderas ofta med frågeformulär men forskning pekar på att funktionella tester som mäter fysisk funktion och aktivitetsmätare som mäter fysisk aktivitet kan komplettera användandet av frågeformulär för att bättre förstå patientens hälsa.

SYFTE. Det övergripande syftet med avhandlingen var att undersöka aspekter av att mäta fysisk funktion och fysisk aktivitet hos patienter med ländryggssmärta.

METOD. I studie I summerades tidigare forskning gällande reliabilitet, validitet och känslighet för förändring hos funktionella tester som används för att mäta fysisk funktion hos patienter med ländryggssmärta. I studie II–IV inkluderades patienter med SRS som skulle genomgå steloperation av ländryggen. I studie II undersöktes känslighet för förändring hos fyra funktionella tester genom att testa fem hypoteser. Vidare undersöktes den minsta kliniskt relevanta förändringen hos testerna med hjälp av den så kallade "optimal cutoff point"-metoden. I studie III undersöktes patienternas preoperativa fysiska aktivitetsnivå med hjälp av en aktivitetsmätare. Vidare användes regressionsanalys för att undersöka hur patienternas preoperativa fysiska aktivitetsnivå var relaterad till tänkbara hinder för att vara mer fysiskt aktiv, såsom rörelserädsla, katastrof-

tankar och depression. I studie IV användes regressionsanalys för att studera om preoperativa variabler kunde förutsäga graden av fysisk aktivitet och funktionshinder sex månader efter steloperation av ländryggen.

RESULTAT. I studie I uppvisade fem-repetitioners sitt-stå test, fem-minuters gångtest, 15-meters gångtest, "progressive isoinertial lifting evaluation" och "timed up-and-go" bäst resultat för reliabilitet och validitet för patienter med långvarig ländryggssmärta. Av dessa tester hade fem-repetitioners sitt-stå test även positiva resultat för känslighet för förändring. En-minuts trapptest uppvisade positiva resultat för både reliabilitet och känslighet för förändring. I studie II–IV inkluderades 118 patienter med SRS. I studie II uppvisade en-minuts trapptest, 15-meters gångtest och "timed up-and-go" positiva resultat för känslighet för förändring. I studie III uppnådde 98 patienter inte WHO:s rekommendationer för fysisk aktivitet, av vilka 32 inte åstadkom en enda minut av de 150 minuterna av fysisk aktivitet per vecka som krävs för att uppnå rekommendationerna. Den låga fysiska aktivitetsnivån var relaterad till graden av funktionshinder och rörelserädsla. I studie IV var en låg preoperativ fysisk aktivitetsnivå och en hög tilltro till sin egen förmåga att träna signifikanta prediktorer till förändringen av fysisk aktivitetsnivå efter steloperation av ländryggen. Vidare var en hög preoperativ nivå av funktionshinder och låga preoperativa nivåer av katastroftankar och tilltro till egen förmåga att träna signifikanta prediktorer till förändringen av graden av funktionshinder efter operationen.

KONKLUSION. Femton-meters gångtest och "timed up-and-go" uppvisade bra resultat för reliabilitet, validitet och känslighet för

förändring och är rekommenderade funktionella tester för att mäta fysisk funktion hos patienter med SRS som genomgår steloperation av ländryggen. Möjligtvis kan framtida pre- och postoperativa interventioner som riktar in sig på lågaktiva patienters rörelserädsla och grad av funktionshinder öka patienternas fysiska aktivitetsnivå. Prediktionsmodellen för fysisk aktivitet kan möjligen användas i kliniken för att förutsäga vilka patienter som är i behov av extra pre- och postoperativa interventioner för att nå en högre fysisk aktivitetsnivå.

LIST OF PAPERS

This thesis is based on the following studies, which are referred to in the text by their Roman numerals.

I.   Jakobsson, M., Gutke, A., Mokkink, L., Smeets, R., Lundberg M. (2018). Level of Evidence for Reliability, Validity, and Responsiveness of Physical Capacity Tasks Designed to Assess Functioning in Patients with Low Back Pain: A Systematic Review Using the COSMIN Standards. *Physical Therapy Journal.* doi:10.1093/ptj/pzy159

II.   Jakobsson, M., Brisby, H., Gutke, A., Lundberg M., Smeets, R. One-Minute Stair Climbing, 50-Foot Walk, and Timed Up-and-Go Were Responsive Measures for Patients with Chronic Low Back Pain Undergoing Lumbar Fusion Surgery. Submitted.

III. Lotzke, H., Jakobsson, M., Gutke, A., Hagströmer, M., Brisby, H., Hägg, O., Smeets, R., Lundberg, M. (2018). Patients with Severe Low Back Pain Exhibit a Low Level of Physical Activity Before Lumbar Fusion Surgery: A Cross-Sectional Study. *BMC Musculoskeletal Disorders*, 19(1):365. doi:10.1186/s12891-018-2274-5

IV.   Jakobsson, M., Brisby, H., Gutke, A., Hagg, O., Lotzke, H., Smeets, R., Lundberg, M. (2018). Prediction of Objectively Measured Physical Activity and Self-Reported Disability Following Lumbar Fusion Surgery. *World Neurosurgery*, 121:e77-e88. doi:10.1016/j.wneu.2018.08.229

CONTENTS

ABBREVIATIONS

DDD	Degenerative disc disease
GPE	Global perceived effect
HADS	Hospital Anxiety and Depression Scale
ICD-11	International Classification of Diseases, 11th Revision
ICF	International Classification of Functioning, Disability and Health
LBP	Low back pain
MIC	Minimal important change
ODI	Oswestry Disability Index
PCS	Pain Catastrophizing Scale
PROM	Patient-reported outcome measure
PROSPERO	International Prospective Register of Systematic Reviews
RMDQ	Roland-Morris Disability Questionnaire
ROC	Receiver operating characteristic
SEES	Self-Efficacy for Exercise Scale
Swespine	Swedish National Quality Registry for Spine Surgery
TSK	Tampa Scale for Kinesiophobia
WHO	World Health Organization

DEFINITIONS

Capacity	The ability to execute a task or an action in a standardized environment (World Health Organization, 2001)
Construct validity	The degree to which the scores of a measurement instrument are consistent with hypotheses based on the assumption that the measurement instrument validly measures the construct to be measured (Mokkink et al., 2010a)
Content validity	The degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured (Mokkink et al., 2010a)
Criterion validity	The degree to which the scores of a measurement instrument are an adequate reflection of a 'gold standard' (Mokkink et al., 2010a)
Exercise	A subset of physical activity that is planned, structured, and repetitive and has as a final or an intermediate objective the improvement or maintenance of physical fitness (Caspersen et al., 1985)
Face validity	The degree to which a measurement instrument indeed looks as though it is an adequate reflection of the construct to be measured (Mokkink et al., 2010a)
Fear of movement	A specific fear of movement and physical activity that is (wrongfully) assumed to cause reinjury (J. W. S. Vlaeyen et al., 1995)
Hypothesis testing	The degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured (Mokkink et al., 2010a)
Interpretability	The degree to which one can assign qualitative meaning - that is, clinical or commonly understood connotations – to a measurement instrument's quantitative scores or change in scores (Mokkink et al., 2010a)
Kinesiophobia	An excessive, irrational, and debilitating fear of physical movement and activity resulting from a feeling of vulnerability to painful injury or reinjury (Kori et al., 1990)
Low back pain	Pain and discomfort, localized below the costal margin and above the inferior gluteal folds, with or without referred leg pain (van Tulder et al., 2006)
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured (Mokkink et al., 2010a)
Minimal important change	The smallest change score that patients perceive as important (Mokkink et al., 2010a)
Pain catastrophizing	An exaggerated negative mental set brought to bear during actual or anticipated painful experience (Sullivan et al., 2001)
Performance	What an individual does in his or her current environment (World Health Organization, 2001)
Physical activity	Any bodily movement produced by skeletal muscle that results in a substantial increase over the resting energy expenditure (Caspersen et al., 1985)
Reliability	The proportion of the total variance in the measurements which is due to 'true' differences between patients (Mokkink et al., 2010a)
Responsiveness	The ability of a measurement instrument to detect change over time in the construct to be measured (Mokkink et al., 2010a)
Self-efficacy	The conviction that one can successfully execute the behavior required to produce the outcomes (Bandura, 1977)
Smallest detectable change	The smallest change that can be detected by the measurement instrument, beyond measurement error (Mokkink et al., 2010a)
Validity	The degree to which a measurement instrument measures the construct(s) it purports to measure (Mokkink et al., 2010a)

INTRODUCTION

1.1 LOW BACK PAIN

Low back pain (LBP) has been defined as "pain and discomfort, localized below the costal margin and above the inferior gluteal folds, with or without referred leg pain" (van Tulder et al., 2006). LBP is often categorized according to the duration of symptoms, with acute pain being of < 6 weeks duration, sub-acute pain being of 6–12 weeks duration, and chronic pain being of ≥ 12 weeks duration (Loeser et al., 2011).

LBP is common in people of all ages (Hartvigsen et al., 2003; Hoy et al., 2012; Steve J. Kamper et al., 2016), and a systematic review of 65 prevalence studies reported that the median lifetime prevalence of LBP was 42.0% (25th–75th percentile: 15.1–60.4%) (Hoy et al., 2012). This systematic review also showed that LBP is slightly more common in women than in men and most frequently occurs between the ages of 40 and 69 years.

Hoy et al. (2014) studied the global burden of LBP and concluded that LBP appears to cause more years lived with disability than any other condition. This finding suggests that LBP has major personal and financial consequences globally (Hoy et al., 2014). Results of a systematic synthesis of 42 qualitative studies have suggested that common psychosocial problems for individuals with LBP are damaged relationships, psychological problems, problems with meeting social expectations and obligations, and the inability to work or participate in other meaningful activities (Froud et al., 2014).

LBP can have many different causes, e.g. fractures, tumors, or degenerative changes (Maher et al., 2017). However, a specific nociceptive cause of LBP is rarely identified (Hartvigsen et al., 2018; Maher et al., 2017). LBP that is believed to be caused by degenerative changes is commonly referred to as degenerative lumbar conditions that include, for example, lumbar disc herniation, lumbar spinal stenosis, and motion-elicited chronic LBP hypothesized to be caused by degenerative disc disease (hereafter referred to as chronic LBP due to DDD) (Modic et al., 2007).

Chronic LBP due to DDD defines a subgroup of individuals with chronic LBP; it is characterized by a combination of motion-elicited and clinically provocable pain with corresponding (i.e. in the same region) radiological findings of degenerative changes in one or a few of the lumbar intervertebral segments (de Schepper et al., 2010; Fritzell et al., 2001; Modic et al., 2007; Willems et al., 2011). The degenerative changes include disc height reduction, Modic changes of the vertebral endplates, and facet arthrosis in varying combinations (Modic et al., 2007).

1.1.1 Lumbar fusion surgery for patients with chronic low back pain due to degenerative disc disease

Lumbar fusion surgery is a treatment option for patients with chronic LBP due to DDD, and it is usually considered first after non-surgical interventions have proven to be unsuccessful (Brox et al., 2003; J. Fairbank et al., 2005; Fritzell et al., 2001; Phillips et al., 2013; Willems et al., 2011). The rationale for lumbar fusion surgery is that pain originating from an intervertebral segment during movement of the spine can be alleviated by

restricting the movement of that segment by fixation (Phillips et al., 2013). Lumbar fusion surgery is most often combined with a postoperative rehabilitation program (Gilmore et al., 2015; Madera et al., 2017) and sometimes with prehabilitation programs that are used before surgery, with the aim of optimizing postoperative outcomes (Cabilan et al., 2016; Gilmore et al., 2015).

Over the past two decades, the number of lumbar fusion operations has constantly increased worldwide, including the USA, the United Kingdom, and Sweden (Deyo et al., 2005; Fritzell et al., 2018; Kalakoti et al., 2016; Rajaee et al., 2012; Strömqvist et al., 2013; The Health and Social Care Information Centre, 2016). In 2011, lumbar fusion surgery caused the highest aggregate hospital costs of any surgical procedure in the USA (Weiss et al., 2014). According to the Swedish National Quality Registry for Spine Sur-

gery (Swespine), approximately 600 patients per year undergo lumbar fusion surgery for chronic LBP due to DDD in Sweden (Fritzell et al., 2018).

The mean age of the patients with chronic LBP due to DDD who undergo lumbar fusion surgery in Sweden is 46 years, so most of them have many years left in the workforce (Fritzell et al., 2018). The patients often describe their symptoms as being dull pain in the lower back that is aggravated by increased mechanical loading and certain movements of the spine (Modic et al., 2007; Willems et al., 2011). Patients with chronic LBP due to DDD have a higher preoperative level of back pain intensity on average than the other patient groups registered in Swespine who undergo elective lumbar spine surgery (e.g. patients with lumbar disc herniation and lumbar spinal stenosis) (Fritzell et al., 2018).

1.2 LOW BACK PAIN AND HEALTH

Regardless of the cause of LBP, the condition can have a significant impact on a patient's health (Froud et al., 2014; Hoy et al., 2014; Shiri et al., 2010; Von Korff et al., 2005). The ultimate goal of both conservative and surgical interventions for patients with LBP is therefore to improve the patients' health (Bernstein et al., 2017; Qaseem et al., 2017). But what is health, and how can it be classified? There are many classifications of health, but for the purpose of this thesis, the International Classification of Functioning, Disability, and Health (ICF) (World Health Organization, 2001) will be used.

1.2.1 The International Classification of Functioning, Disability, and Health

The ICF is a classification system for health with a biopsychosocial perspective. In contrast to the biomedical perspective of health,

the biopsychosocial perspective incorporates psychological and social factors in addition to biological factors (Engel, 1977; Gatchel et al., 2007; Waddell, 1992). The ICF was designed to standardize the terminology and measurement of health to facilitate collaboration between different health professionals and between different countries (World Health Organization, 2001). Information on health provided by the ICF together with information on diagnoses described in the International Classification of Diseases, 11th Revision (ICD-11) is thought to give a more comprehensive picture of an individual's health than when using one of the classification systems alone (World Health Organization, 2001). In the ICF, a person's health is determined by an interaction between his/her health condition (such as LBP) and his/her functioning and disability.

Functioning is divided into three domains (Figure 1):

- *Body functions and structures* refer to psychological and physiological processes and their anatomical structures, such as pain, range of motion, and muscle strength of the lower back.

- *Activity* refers to the person's ability to perform tasks in his/her daily life, such as the ability to walk, lift, and rise up from a chair.

- *Participation* describes the person's involvement in a life situation, such as the ability to work, to socialize with friends, or to buy groceries.

Measuring aspects of the three domains of functioning gives the healthcare professional information about "neutral" or "positive" aspects of a patient's health. Disability is closely related to functioning, but instead concerns "negative" aspects of health, i.e., impairments, activity limitations, and participation restrictions. By measuring several aspects of the domains of functioning and disability, the healthcare professional can gain a comprehensive overview of the patient's health (World Health Organization, 2001b). The measurement of functioning and disability is described in detail in sections 1.3.2 and 1.3.3.

Furthermore, personal factors (e.g. gender, age, and coping strategies) and environmental factors (e.g. family, work, and education level) can work as facilitators of or barriers to a patient's health (World Health Organization, 2001). Potential barriers to health for patients with LBP are described in section 1.3.5.

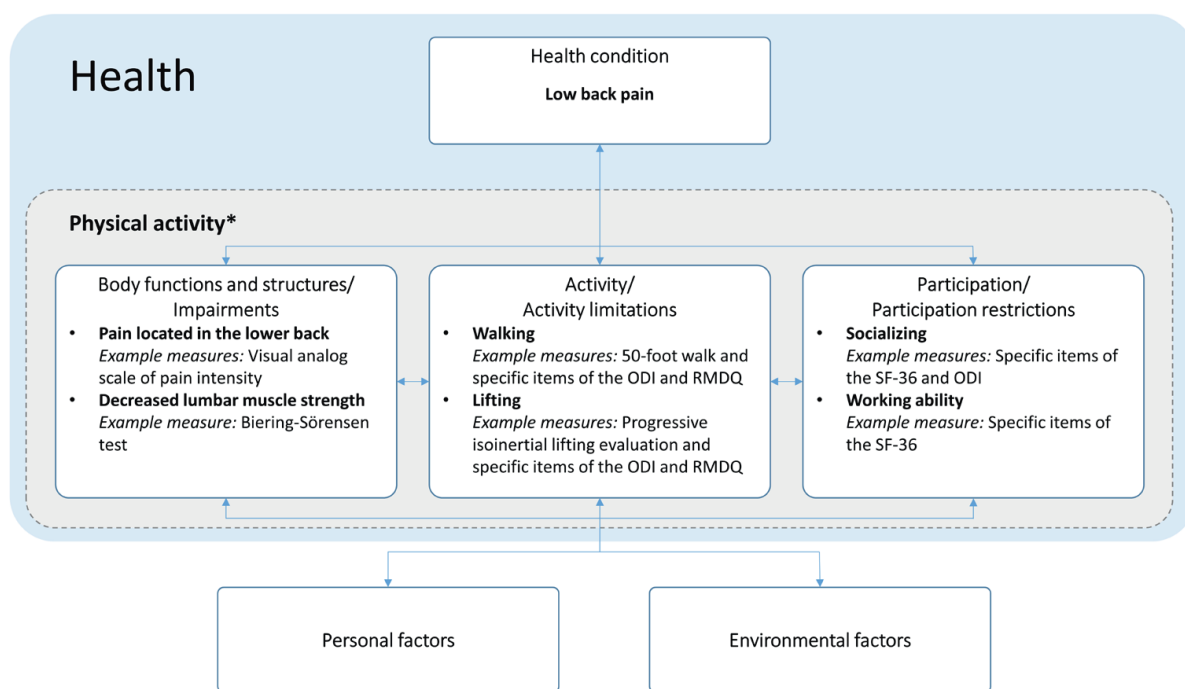**Figure 1.** Overview of the components of the International Classification of Functioning, Disability, and Health, exemplified for a hypothetical patient with low back pain.

*Physical activity is not incorporated in the original ICF model. The current model is a modified version of that of van der Ploeg et al. (2004).

ODI, Oswestry Disability Index; RMDQ, Roland-Morris Disability Questionnaire; SF-36, Short-Form (36) health survey.

1.2.2 Physical activity and health

Physical activity is another important concept in relation to health, although it is not explicitly incorporated in the ICF. Physical activity has been defined as "any bodily movement produced by skeletal muscle that results in a substantial increase over the resting energy expenditure" (Caspersen et al., 1985). Physical activity has been considered to have an effect on many aspects of the domains of the ICF (Figure 1) (van der Ploeg et al., 2004). For instance, an increased level (duration, intensity, and frequency) of physical activity can improve cardiopulmonary function and muscle strength in the body function and structures domain. In the activity domain, an increased level of physical activity may lead to better walking or lifting ability. In the participation domain, an increased level of physical activity can result in increased working ability (van der Ploeg et al., 2004).

There is a dose-response relationship between the level of physical activity and positive health benefits such as a reduced risk of non-communicable diseases such as diabetes, cancer, and cardiovascular disease (I. M. Lee et al., 2012; Wen et al., 2011; World Health Organization, 2009a). This relationship means that people with the lowest level of physical activity can gain the greatest effects on health by increasing their level of physical activity (Wen et al., 2011). Estimates suggest that if sedentary individuals were to increase their physical activity level sufficiently, 3.2–5.3 million deaths could be prevented annually (I. M. Lee et al., 2012; World Health Organization, 2009a). Thus, from a public health point of view, it is particularly important to reach those who have a low level of physical activity (World Health Organization, 2009a).

In the context of patients with LBP, physical activity is usually described as a component of LBP interventions to reduce disability and improve functioning (Airaksinen et al., 2006; Bernstein et al., 2017; Qaseem et al., 2017). However, in this thesis, physical activity is viewed from a broader health standpoint in that physical activity can also reduce the risk of non-communicable diseases (I. M. Lee et al., 2012; Wen et al., 2011; World Health Organization, 2009a).

The authors of a recently published call for action advocated a stronger emphasis on health in the interventions of patients with LBP, such that interventions would be aligned with the World Health Organization (WHO) action plans to improve health and prevent non-communicable diseases (Buchbinder et al., 2018). This also includes the WHO global recommendations on physical activity for health (World Health Organization, 2009b):

1. "Adults aged 18–64 should do at least 150 minutes of moderate-intensity aerobic physical activity throughout the week or do at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week or an equivalent combination of moderate- and vigorous-intensity activity.

2. Aerobic activity should be performed in bouts of at least ten minutes duration.

3. For additional health benefits, adults should increase their moderate-intensity aerobic physical activity to 300 minutes per week, or engage in 150 minutes of vigorous-intensity aerobic physical activity per week, or an equivalent combination of moderate- and vigorous-intensity activity.

4. Muscle-strengthening activities should be done involving major muscle groups on two or more days a week."

This thesis concerns points 1 and 2 of the recommendations. Point 1 of the recommendations is an indication of the fact that moderate-intensity and vigorous-intensity physical activity results in greater health benefits than lower intensities (World Health Organization, 2009b). Moderate-intensity physical activity can be compared to a brisk walk (~4.0 km/h) and vigorous-intensity physical activity is equivalent to running (~6.5 km/h) (Ainsworth et al., 2011).

It is often assumed that patients with LBP who report having a high level of disability will be less physically active than patients with no back problems (Lin et al., 2011). If LBP does indeed cause patients to be less physically active, this would mean that LBP can increase the risk of negative health effects, considering that the patients might not reach a health-enhancing level of physical activity. However, research on physical activity in patients with lumbar degenerative conditions who undergo lumbar spine surgery is scarce. Results of previous research on the level of physical activity in these patients are presented in section 1.3.4.

1.3 MEASUREMENT OF HEALTH IN PATIENTS WITH LOW BACK PAIN

So far, this introduction has described that LBP is a possible threat to a patient's health in terms of disability, reduced functioning, and a lower level of physical activity. But how can health be measured in patients with LBP?

Up until the 1980s, the assessment of the outcome of lumbar spine surgery was to a great extent assessed from the surgeon's point of view (Deyo, 1988). For example, scales were used in which the technical success of the surgical procedure was scored "excellent," "good," "moderate," or "bad" (Getty, 1980). During the 1980s and the early 1990s, several patient-reported outcome measures (PROMs) were developed (EuroQoL Group, 1990; J. C. T. Fairbank et al., 1980; Roland et al., 1983; Ware et al., 1992). As the term suggests, PROMs are designed to measure the patient's view of his/her health rather than the clinician's view (Food and Drug Administration, 2009). Examples of PROMs used for patients with LBP are PROMs concerning pain (e.g., visual analog scales and numerical rating scales), disability (e.g., the Oswestry Disability Index and the Roland-Morris Disability Questionnaire), health-related quality of life (e.g. the 36-item Short-Form and EuroQoL-5D), and global assessment scales (EuroQoL Group, 1990; J. C. T. Fairbank et al., 1980; Fischer et al., 1999; Roland et al., 1983; Ware et al., 1992).

PROMs are still among the most commonly used measures in the assessment of both conservative and surgical interventions for LBP (Chapman et al., 2011; Chiarotto et al., 2016), but other approaches for measuring health have emerged during the last three decades. In the 1990s, researchers developed and suggested the use of so-called physical capacity tasks, in which the patient's functioning was assessed by having him/her perform a standardized activity in the clinical setting (Harding et al., 1994; Simmonds et al., 1998). At the beginning of the 2000s, assessment of physical activity in patients with LBP became increasingly common, in part due to the development of portable activity monitors such as pedometers and accelerometers (van Weering et al., 2009; Verbunt et al., 2001).

In summary, there are several different ways of measuring health in patients with LBP. In this thesis, I will focus mainly on physical capacity tasks that measure functioning,

and accelerometers that measure the level of physical activity. However, before going into more detail about these outcome measures, I will first briefly cover reliability, validity, and responsiveness—as it is essential to consider these measurement properties in the measurement of health.

1.3.1 Reliability, validity, responsiveness, and interpretability of health outcome measures

It is important that outcome measures in clinical work and research have sufficient evidence for reliability, validity, and responsiveness. If the evidence is insufficient, there is a significant risk of getting imprecise or biased results in the evaluation of health interventions (Brakenhoff et al., 2018a; Brakenhoff et al., 2018b; de Vet et al., 2011; Streiner et al., 2008).

Reliability has been defined as "the degree to which the measurement instrument is free from measurement error" (Mokkink et al., 2010a). Measurements performed on two or more occasions in the same individual may yield different results due to many factors such as biological variability, the mood of the person, or characteristics of the outcome measure itself (Streiner et al., 2008). Reliability concerns how the variability of individuals is related to the measurement error and indicates how well participants can be distinguished from each other despite this measurement error (de Vet et al., 2011). *Measurement error has been* defined as "the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured" (Mokkink et al., 2010a). If the measurement error is large, small changes in a patient cannot be differentiated from measurement error. Researcher and clinicians therefore need to know the extent of this error in order to interpret the change in a patient correctly (de Vet et al., 2006).

Validity has been defined as "the degree to which a measurement instrument measures the construct(s) it purports to measure" (Mokkink et al., 2010a). Many constructs in the health sciences are readily observable physical quantities such as weight, blood glucose, or body temperature. For such constructs, there is usually a criterion measure (or 'gold standard') against which new outcome measures can be compared. If the correlation between the new measure and the criterion measure is high, the validity of the new outcome measure can be considered to be adequate. This way of assessing validity is usually referred to as *criterion validity* (Streiner et al., 2008). However, criterion validity is often not possible to assess for abstract constructs such as disability and functioning, as there are rarely suitable criteria to compare against. For such constructs, the assessment of validity relies on posing and testing hypotheses based on the knowledge of the construct of interest. This way of assessing validity is referred to as *construct validity* (de Vet et al., 2011; Streiner et al., 2008). *Content validity* refers instead to the relevance and comprehensiveness of an outcome measure. Relevance is about whether the items of the outcome measure appropriately reflect the construct of interest. Comprehensiveness denotes the degree to which all aspects of the construct are covered by items in the outcome measure (de Vet et al., 2011). *Face validity* is a subcategory of content validity, and reflects a subjective view of whether the measurement instrument "looks as if" it measures what it is designed to measure. Face validity is therefore considered to be a less strict form of content validity (Mokkink et al., 2010a).

Responsiveness has been defined as "the ability of a measurement instrument to detect change over time in the construct to be measured" (Mokkink et al., 2010a). Responsiveness is closely related to validity, but concerns change scores instead of scores collected at one time point (Streiner et al., 2008). As re-

sponsiveness involves change over time, it is an important measurement property to consider when measuring the effect of a health intervention (de Vet et al., 2011).

Interpretability has been defined as "the degree to which one can assign qualitative meaning—that is, clinical or commonly understood connotations—to a measurement instrument's quantitative scores or change in scores" (Mokkink et al., 2010a). Put simply, interpretability parameters help clinicians and researchers to understand what scores and change scores of an outcome measure really mean (Streiner et al., 2008). One of the most common interpretability parameters is the minimal important change. Minimal important change is used to interpret whether the change scores of an outcome measure are of importance to patients and not only statistically significant (de Vet et al., 2006).

It is important to note that reliability, validity, responsiveness, and interpretability are not inherent properties of an outcome measure. Instead, these measurement properties depend on an interaction between the outcome measure itself, the individuals who are measured, and the context of the individuals (de Vet et al., 2011; Streiner et al., 2008). Consequently, it is not correct for researchers in a reliability study to conclude that an outcome measure is "reliable." It is instead more correct to state that the outcome measure has been found to be reliable for a particular group of patients in a certain context. Thus, the reliability, validity, and responsiveness of health outcome measures for patients with chronic LBP should, preferably, be investigated specifically in patients with chronic LBP.

1.3.2 Measurement of disability with patient-reported outcome measures

Two of the most commonly used PROMs that measure disability in patients with LBP are the Oswestry Disability Index (ODI) (J. C. T.

Fairbank et al., 1980) and the Roland-Morris Disability Questionnaire (RMDQ) (Roland et al., 1983). These PROMs were recently recommended in a recently-developed core set of outcome measures for clinical trials of patients with LBP (Chiarotto et al., 2018).

Relating RMDQ and ODI to the ICF framework, the majority of the items of the PROMs measure how the patient perceives that his/her back pain affects common activities, such as walking, lifting, and sitting (J. C. T. Fairbank et al., 1980; Roland et al., 1983). The ODI and RMDQ have therefore been interpreted to mainly measure disability in the activity limitation domain of the ICF (Grotle et al., 2005; Smeets et al., 2007). The PROMs have also been interpreted as being in line with the ICF concept of performance (Grotle et al., 2005; Smeets et al., 2007; Wittink, 2005), which denotes what activities a person does in his/her usual environment (World Health Organization, 2001).

A strength of PROMs is that they give the healthcare professional an indication of how the patient perceives his/her health (Streiner et al., 2008). Researchers have also recognized the added value of PROMs in improving the communication between caregivers and patients, and for detecting health problems that would otherwise have gone unnoticed (Valderas et al., 2008). Moreover, PROMs are time-efficient and do not require advanced instruments or high administration costs (de Vet et al., 2011).

However, previous research has raised concerns regarding disability PROMs such as low- to very low-quality evidence for content validity (Chiarotto et al., 2017) and floor and ceiling effects (Brodke et al., 2017; Lauridsen et al., 2006; Pekkanen et al., 2011). Moreover, disability PROMs are not suitable for all individuals, as they may be too challenging regarding the cognitive and language

skills needed to fill them out (Gautschi et al., 2016c; Guralnik et al., 1989). Furthermore, while the subjective nature of PROMs is a strength for obtaining the patient's view of his/her health, it may also constrain comparison between patients (Gautschi et al., 2014; Staartjes et al., 2018). Moreover, clinical experience and scientific evidence also suggest that there are often discrepancies between how patients score PROMs and how they actually move and perform activities when they are observed by others (Gautschi et al., 2016c; C. E. Lee et al., 2001; Simmonds et al., 1998).

1.3.3 Measurement of functioning with physical capacity tasks

Several researchers have recommended the use of so-called physical capacity tasks in addition to disability PROMs (Gautschi et al., 2014; Harding et al., 1994; Jespersen et al., 2018; Simmonds et al., 1998; Smeets et al., 2006; Wittink, 2005). In a physical capacity task, the patient performs a standardized activity in the clinic instead of self-reporting his/her ability to perform the activity (Harding et al., 1994; Simmonds et al., 1998). The activities assessed in physical capacity tasks usually involve those that are commonly affected by LBP, such as walking or lifting (Simmonds et al., 1998; Smeets et al., 2006). An example of a physical capacity task is the timed up-and-go, which measures the time it takes for a person to rise from a chair, walk three meters, turn around, walk back to the chair and sit down (Simmonds, 1998). In the context of lumbar spine surgery, physical capacity tasks have mostly been used for patients with lumbar spinal stenosis (Deen et al., 2000; Jespersen et al., 2018; Pratt et al., 2002). However, in recent years, the tests have received more attention for other diagnoses, including chronic LBP due to DDD (Gautschi et al., 2014; Staartjes et al., 2018).

Physical capacity tasks have been given different labels by different researchers, such as "physical performance tests" (Simmonds et al., 1998) and "functional assessments tests" (Wittink, 2005), but in this thesis, they will be referred to as physical capacity tasks to be in line with the ICF concept of capacity. Capacity is defined as the "highest possible level of functioning of a person in a given domain at a given moment, measured in a standardized environment" (World Health Organization, 2001). By using outcome measures that measure capacity (e.g. physical capacity tasks) in addition to those that measure performance (e.g. disability PROMs), the ICF suggests that the healthcare professional can acquire a comprehensive overview of a patient's health status (World Health Organization, 2001). There is also empirical evidence supporting the idea that physical capacity tasks and disability PROMs indeed measure different aspects of a patient's health (Conway et al., 2011; Gautschi et al., 2016b; C. E. Lee et al., 2001).

Research also suggests that physical capacity tasks have a number of benefits over disability PROMs. First, physical capacity tasks are suggested to be relatively uninfluenced by education level, language, and cognitive skills (Guralnik et al., 1989; Simmonds et al., 1998; Teixeira Da Cunha-Filho et al., 2010; Wand et al., 2010). Second, they are usually not associated with the floor and ceiling effects often seen in disability PROMs (Brodke et al., 2017; Lauridsen et al., 2006; Pekkanen et al., 2011). Third, a recent study showed that patients with LBP who had undergone lumbar spine surgery were more than six times as likely to prefer performing a physical capacity task (timed up-and-go) than completing a set of PROMs (Joswig et al., 2017).

Simmonds et al. (1998) performed one of the first studies to investigate the measurement properties of physical capacity tasks

specifically for patients with chronic LBP. The study showed that many of the physical capacity tasks had support for adequate reliability and validity. Several studies with similar results followed, such as Pratt et al. (2002), Magnussen et al. (2004), and Smeets et al. (2006).

> *Identified research gap #1:* To the best of my knowledge, no study has made a synthesis of previous findings of reliability, validity, and responsiveness of physical capacity tasks that measure functioning in patients with LBP.

1.3.4 Measurement of physical activity with accelerometers

Physical activity can be described and quantified in terms of four principal dimensions: type, frequency, duration, and intensity (Strath et al., 2013). Type denotes the activity performed, e.g. walking, running, or lifting. Type can also refer to the biomechanical or physiological demands, such as strength training, plyometric training, and aerobic or anaerobic activities. Frequency describes the number of sessions of an activity in a given time period. Duration is simply the length of the activity. Intensity denotes the energy expenditure or approximate effort in performing an activity (Welk, 2002).

Measurement of the level (frequency, duration, and intensity) of physical activity can be achieved through, for example, PROMs (e.g. the Baecke physical activity questionnaire) or wearable monitors (e.g. pedometers and accelerometers) (Welk, 2002). Accelerometers are motion sensors that measure body movement from changes in velocity in relation to time. The raw data of the accelerometer can then be transformed into the time spent per day with different intensities of physical activity, usually light-, moderate-, or vigorous-intensity physical activity (Troiano et al., 2008). Accelerometers are recommended over self-reports, since they are not reliant on accurate recall of the frequency, duration, and intensity of physical activity and are less subject to overestimations and social desirability (Cerin et al., 2016; Prince et al., 2008; Slootmaker et al., 2009). Accelerometers also provide advantages compared to pedometers because they measure several aspects of physical activity such as duration and intensity, and not just the number of steps (Chen et al., 2005).

The research on physical activity in patients with lumbar degenerative conditions who undergo lumbar spine surgery is scarce (Lindbäck et al., 2017; Mobbs et al., 2016; Norden et al., 2017; Rolving et al., 2013; Smuck et al., 2018). Of the studies that used accelerometers, one is difficult to draw conclusions from as it had a small sample size (n = 30) and used an accelerometer with limited support for validity (Mobbs et al., 2016). Another study showed that 4% of patients with lumbar spinal stenosis scheduled for decompression surgery fulfilled the WHO recommendations on physical activity, which indicates a very low level of physical activity (Norden et al., 2017). A follow-up study that included the same patients did not show any statistically significant change in the patients' level of physical activity after surgery (Smuck et al., 2018). These results can, however, not be extrapolated to patients with LBP due to DDD who undergo lumbar fusion surgery. First, patients with LBP due to DDD who are scheduled for lumbar fusion surgery are on average younger, do not have severe leg symptoms such as neurogenic claudication, and can walk longer distances than patients with lumbar spinal stenosis (Fritzell et al., 2018; Strömqvist et al., 2013). Second, the previous studies that used accelerometers were performed in a non-European context, which may also affect the level of physical activity (Hagströmer et al., 2010).

1.3.5 Measurement of fear-avoidance variables to identify barriers to and predictors of health

Improved health is usually the primary goal of interventions for patients with LBP (Bernstein et al., 2017; Qaseem et al., 2017), and I explained in the previous sections that high-quality outcome measures are needed to evaluate the effectiveness of health interventions. Previous research suggests that the effectiveness of such interventions can be increased by targeting specific barriers to health in patients with LBP (J. W. Vlaeyen et al., 2012; J. W. S. Vlaeyen et al., 1995; Woby et al., 2007).

The ICF states that personal and environmental factors can work as barriers to a patient's health (World Health Organization, 2001). High age and high BMI are examples of personal factors that can act as such barriers (Bauman et al., 2012). A lack of social support and restricted access to exercise facilities are examples of environmental barriers (World Health Organization, 2001). For the purposes of this thesis, variables in the cognitive behavioral fear-avoidance model developed by Vlaeyen et al. (J. W. S. Vlaeyen et al., 1995) and modified by Woby et al. (2007) and Lotzke et al. (2016) are considered to be potential barriers to the health components functioning, disability, and physical activity.

The cognitive behavioral fear-avoidance model by Vlaeyen et al. (1995) was developed from a biopsychosocial perspective. The version of the model in this thesis is the modification presented by Lotzke et al. (2016) (hereafter referred to as "the modi-fied fear-avoidance model"). The modified fear-avoidance model includes self-efficacy for exercise and physical activity in addition to the variables in the original cognitive behavioral fear-avoidance model. The modified fear-avoidance model describes two possible trajectories, depending on how a patient interprets a pain episode. In the trajectory pictured to the right in Figure 2, patients who view the pain as non-threatening and transitory will go back to their usual activities and experience a gradually reduced level of disability and depression, and a higher level of physical activity. In the other trajectory, the model suggests that patients who respond to pain with catastrophizing thoughts, such as incorrectly interpreting the pain as a sign of serious injury, may develop a fear of movement (kinesiophobia in its extreme form). The model also proposes that the fear of movement gradually results in avoidance behavior regarding activities. If the avoidance behavior persists, the model suggests that it will lead to disability, depression, and a lower level of physical activity (Lotzke et al., 2016). In line with Woby et al. (2007), the model suggests that individuals with low self-efficacy are more likely to develop avoidance behaviors, disability, and depression.

The fear-avoidance variables have been investigated extensively in patients with chronic LBP in a non-surgical context (Denison et al., 2004; Glombiewski et al., 2018; Peters et al., 2005; Pincus et al., 2002; Turner et al., 2000; J. W. Vlaeyen et al., 2012; J. W. S. Vlaeyen et al., 1995; Woby et al., 2007), but there has been less research in patients who undergo lumbar spine surgery.
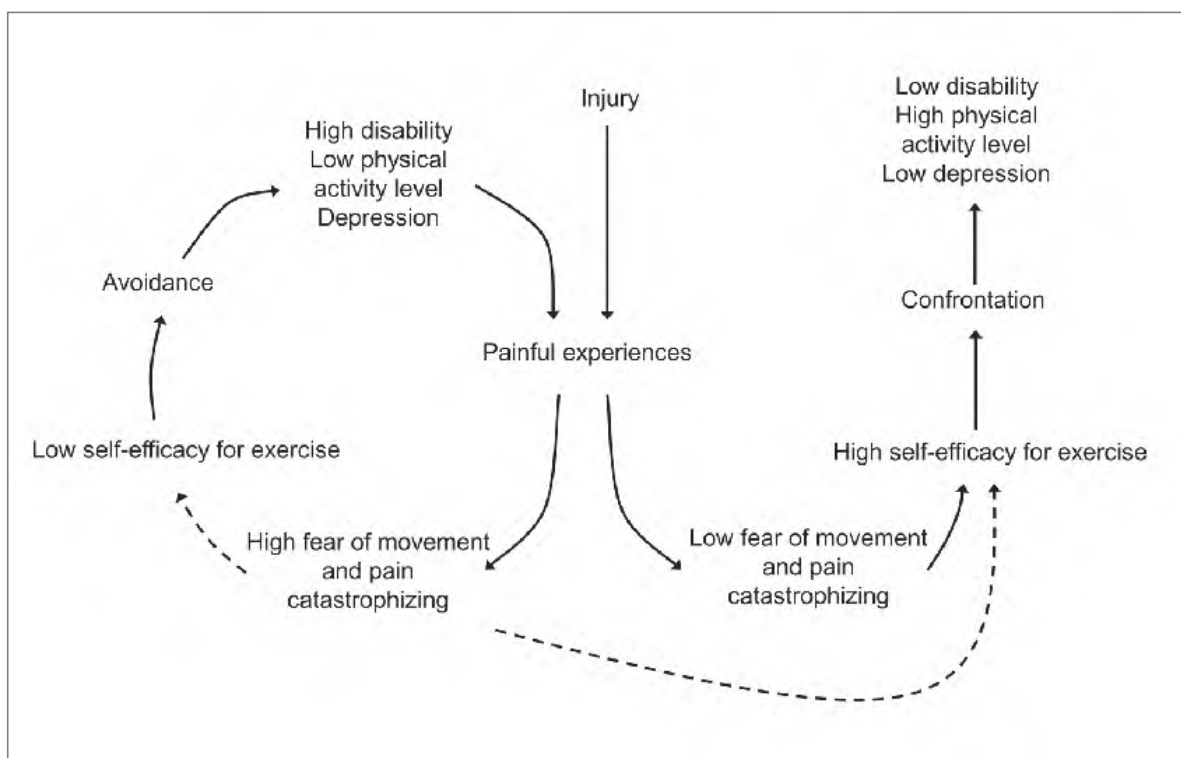
Figure 2. The modified cognitive fear-avoidance model by Lotzke et al. (2016), based on the models of Vlaeyen et al. (1995) and Woby et al. (2007)
Reproduced from BMC Musculoskeletal Disorders, CC permission.

A few studies have shown that fear-avoidance variables appear to be barriers to health in patients undergoing lumbar spine surgery (Abbott et al., 2010b; Lundberg et al., 2011), but other studies have not (Grotle et al., 2004b; Johansson et al., 2016). One study showed that preoperative pain, fear of movement, and depression accounted for 67% of the explained variance in disability in patients with lumbar degenerative conditions (Lundberg et al., 2011). In another study, on patients with lumbar degenerative conditions who were scheduled for lumbar fusion surgery, fear-avoidance variables accounted for 50% of the explained variance in disability (Abbott et al., 2010b).

Fear-avoidance variables have also been investigated as barriers to being physically active (Carvalho et al., 2017; Elfving et al., 2007; Verbunt et al., 2005). The rationale for doing so was that the avoidance behavior caused by pain catastrophizing and fear of movement could cause a decreased level of physical activity (Verbunt et al., 2005; Verbunt et al., 2010). The empirical evidence for the idea that the fear-avoidance variables would be barriers to physical activity in patients with LBP is, however, limited. One study found that a high level of fear of movement or catastrophizing was significantly associated with a low level of physical activity in patients in a non-surgical context (Elfving et al., 2007), whereas another study did not (Carvalho et al., 2017).

Identified research gap #3: To the best of my knowledge, no previous studies have investigated whether fear-avoidance variables are barriers to physical activity in patients with chronic LBP due to DDD who undergo lumbar fusion surgery.

The fear-avoidance variables have also been used to predict health outcomes following lumbar fusion surgery (Abbott et al., 2011;

DeBerard et al., 2003; den Boer et al., 2006; LaCaille et al., 2005; Trief et al., 2000). The outcome of lumbar fusion surgery for chronic LBP is a topic of much debate (Hedlund et al., 2016; Mannion et al., 2016), and randomized controlled trials have shown conflicting evidence as to whether lumbar fusion surgery is superior to non-surgical interventions (Brox et al., 2003; Bydon et al., 2014; J. Fairbank et al., 2005; Foster et al., 2018; Fritzell et al., 2001; Mannion et al., 2016). It is therefore important to develop prediction models that—already before surgery—can help healthcare professionals to identify individuals who are less likely to have a successful outcome of surgery (Mannion et al., 2006). Prediction models can thereby support clinical decision-making to optimize treatment benefit and cost-effectiveness (Mannion et al., 2006; Steyerberg, 2009).

Traditional prediction models for predicting postoperative outcomes following lumbar fusion surgery include variables such as gender (van Susante et al., 1998, smoking (Andersen et al., 2001; Glassman et al., 2000; LaCaille et al., 2005; Trief et al., 2006), and pain duration (Trief et al., 2000; Woertgen et al., 1999). Modern prediction models usually take a biopsychosocial approach using variables such as work status, social support, and variables found in the cognitive behavioral fear-avoidance model (Wilhelm et al., 2017).

Previous studies investigating the predictive value of fear-avoidance variables have demonstrated that high levels of preoperative fear of movement (den Boer et al., 2006), pain catastrophizing (Abbott et al., 2011), and depression (DeBerard et al., 2003; LaCaille et al., 2005; Trief et al., 2000) can predict a less favorable postoperative outcome after lumbar fusion surgery. Self-efficacy is a fear-avoidance variable that has not been investigated as a predictor of the outcome of lumbar spine surgery. However, self-efficacy has been

shown to be predictive of clinical outcomes in patients undergoing various types of other surgeries such as anterior cruciate ligament reconstruction (Everhart et al., 2015) and hip and knee arthroplasty (van den Akker-Scheek et al., 2007; Wylde et al., 2012).

Most studies on the predictive values of fear-avoidance variables have aimed to predict disability and not physical activity (Abbott et al., 2011; DeBerard et al., 2003; den Boer et al., 2006; LaCaille et al., 2005; Trief et al., 2000; Wilhelm et al., 2017). However, considering the beneficial health effects of physical activity (I. M. Lee et al., 2012; World Health Organization, 2009a), it is important to also identify predictors of the patient's level of physical activity after lumbar fusion surgery.

As the variables in the fear-avoidance model have been hypothesized to affect physical activity and not only disability (Lotzke et al., 2016; Verbunt et al., 2005; Verbunt et al., 2010), they appear to also have the potential ability to predict physical activity following lumbar fusion surgery. Furthermore, walking capacity could also be a possible predictor of postoperative changes in both physical activity and disability. The rationale for this is based on findings in previous prediction studies (Gunzburg et al., 2003; Soriano et al., 2010) and also because there is a high correlation between walking capacity and health status (Blain et al., 2010; Montero-Odasso et al., 2005; Ostir et al., 2007; Tabue-Teguo et al., 2015).

Identified research gap #4: To the best of my knowledge, there has been very little previous research on the predictive value of fear-avoidance variables and walking capacity in predicting the postoperative outcome of physical activity and disability in patients with chronic LBP due to DDD who undergo lumbar fusion surgery.

AIMS

The overall aim of the work described in this thesis was to investigate aspects of the measurement of functioning and physical activity in patients with low back pain. The specific aims were:

I. To systematically review the level of evidence of reliability, validity, and responsiveness of physical capacity tasks that are designed to assess **functioning** in patients with LBP (Study I)

II. To investigate the responsiveness and minimal important change of four physical capacity tasks used to assess **functioning** in patients with chronic LBP due to DDD who undergo lumbar fusion surgery (Study II)

III. To investigate the preoperative level of **physical activity** in patients with chronic LBP due to DDD scheduled for lumbar fusion surgery, and furthermore to investigate whether fear-avoidance variables are associated with this level (Study III)

IV. To investigate the predictive value of preoperative fear-avoidance variables, walking capacity, and traditional predictor variables for prediction of postoperative changes in **physical activity** level and disability six months after lumbar fusion surgery in patients with chronic LBP due to DDD (Study IV)

METHODS

An overview of the methods of the four studies in the thesis is given in Table 1. The methods of Study I are presented in a separate section (Section 3.2) while the methods of Studies II–IV are presented together in the same section (Section 3.3). The rationale for this is that Study I was a systematic review and the methods used in this study therefore differed considerably from those that were used in the other studies.

Table 1. Overview of study designs, recruitment, study populations, type of data, and data analysis

	Study I	Study II	Study III	Study IV
Study design	Systematic review of reliability, validity, and responsiveness	Clinimetric study	Cross-sectional study	Prediction study
Recruitment	N/A	Recruitment from surgical waiting lists for lumbar fusion surgery at one university hospital and two private spine clinics, as part of a randomized controlled trial (Lotzke et al., 2016)		
Study population	25 articles containing patients with low back pain of ≥ 6 weeks duration	118 patients with chronic low back pain due to degenerative disc disease scheduled for lumbar fusion surgery		
Type of data used in data analysis	Results of reliability, validity, and responsiveness found in the included articles	PROMs and physical capacity tasks (baseline and 6-month postoperative data)	PROMs and accelerometers (baseline data)	PROMs, physical capacity tasks, and accelerometers (baseline and 6-month postoperative data)
Data analysis	Best-evidence synthesis to determine the level of evidence for reliability, validity, and responsiveness	Responsiveness analysis and minimal important change analysis	Physical activity analysis and regression analysis	Physical activity analysis and regression analysis

PROMs, patient-reported outcome measures

3.1 ETHICAL APPROVAL

Study I did not need ethical approval, as it was a systematic review. Studies II–IV were approved by the Regional Ethical Review Board of Gothenburg (Dnr.586-11 and amendment T 527-15). All the studies in the thesis adhered to the Code of Ethics of the World Medical Association (Declaration of Helsinki). Ethical considerations are discussed in Section 5.5.

3.2 STUDY I

3.2.1 Protocol and registration

A protocol was registered at the International Prospective Register of Systematic Reviews (PROSPERO), http://www.crd.yor.ac.uk/PROSPERO (registration number: CRD42016042011).

3.2.2 Eligibility criteria

Articles that met the following criteria were included:

- **Target population:** The study population had LBP of ≥ 6 weeks duration and was aged ≥ 18 years (van Tulder et al., 2006). Articles that contained pregnant participants or participants suffering from confirmed rheumatic diseases, fibromyalgia, tumors, infections, osteoporosis, structural deformities (e.g. scoliosis), fractures, or cauda equina syndrome were excluded unless data were presented specifically for patients who adhered to the eligibility criteria.

- **Construct:** The test was a measure of "capacity" of the ICF activity domain, defined as "the ability to execute a task or an action in a standardized environment." (World Health Organization, 2001).

- **Outcome measure:** The test was a physical capacity task, defined as (i) a standardized test that is used for an evaluative purpose and that (ii) is administered by an observer, (iii) includes an activity as classified by the ICF that (iv) is performed in a standardized setting, and (v) requires low-cost and readily available portable equipment. Articles that exclusively investigated test batteries and did not present results for individual physical capacity tasks were excluded. If an article cited an original test manual that could not then be obtained, the test was excluded.

- **Article type:** The article presented original data reporting the reliability (including reliability, measurement error, and internal consistency), validity (including content validity, construct validity, and criterion validity), and responsiveness (Mokkink et al., 2010a).

3.2.3 Classification of reliability, validity, and responsiveness

The Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) taxonomy was used to classify measurement properties (reliability, validity, and responsiveness) found in the articles included (Mokkink et al., 2010a). The COSMIN taxonomy was developed by Mokkink et al. (2010a) through an international Delphi study to clarify and standardize the terminology of measurement properties. Figure 3 shows an overview of the interrelationships between the measurement properties in the COSMIN taxonomy.
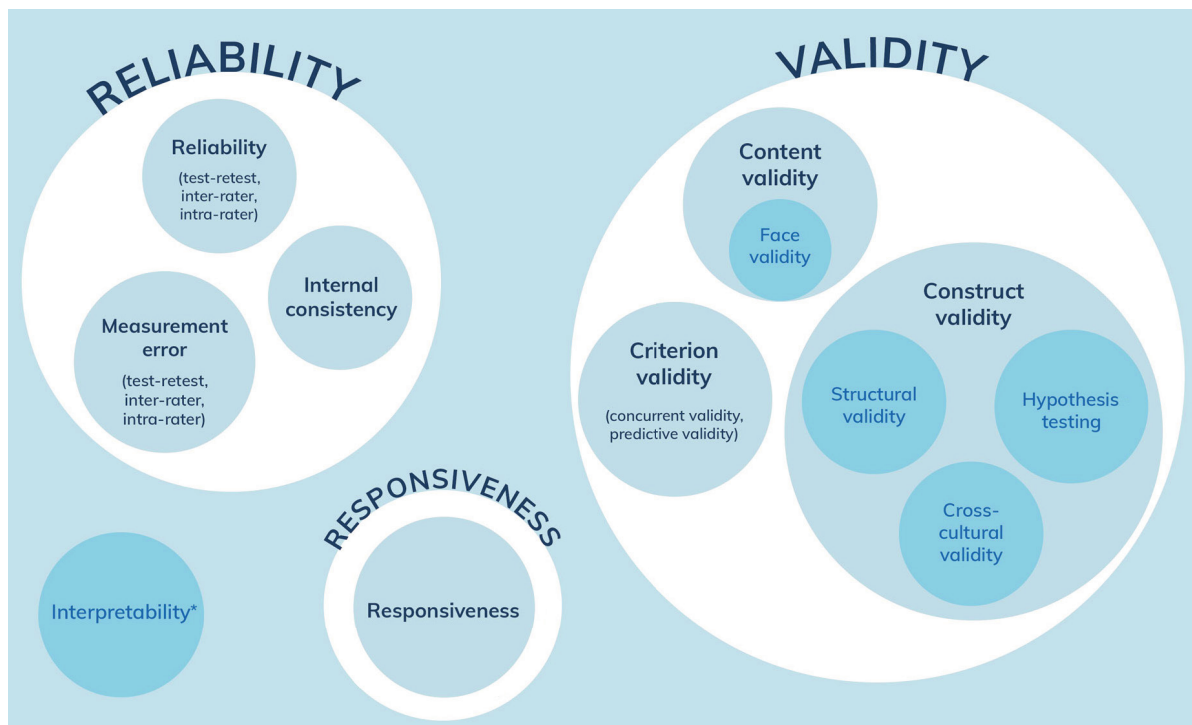
Figure 3. Overview of the interrelationships between measurement properties in the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) taxonomy (Mokkink et al., 2010a).
Reproduced from Journal of Clinical Epidemiology with permission from Elsevier.

*Interpretability is not considered a measurement property in the COSMIN taxonomy because it does not refer to the quality of a measurement instrument. However, interpretability was considered to be sufficiently important to be included in the COSMIN taxonomy.

3.2.4 Information sources

Electronic information sources were MEDLINE (through the interface of Ovid), CINAHL (EBSCOhost), PsycINFO (Ovid), Scopus (Elsevier), and the Cochrane Library (Wiley). Additional information sources were the reference lists of articles from the electronic database search that were identified for full-text reading. We developed a search strategy in collaboration with two medical librarians working at a university library, who then performed the search. The search strategy included three main filters, specifically tailored for each electronic database: the construct of interest, the target population, and measurement properties. No restrictions were applied regarding language.

The last search was performed on August 29, 2018.

3.2.5 Study selection

Two authors screened—independently of each other—titles and abstracts of studies from the database search and hand search. A consensus was reached on which articles should be reviewed in full-text. If consensus could not be reached, a third author was consulted to resolve the disagreement. Subsequently, two authors independently reviewed the full-text articles for eligibility. A third author was consulted if consensus could not be reached regarding the inclusion of full-text articles.

3.2.6 Data collection

One author developed a data extraction form, which was then piloted on five randomly-selected included studies, by three authors. The data extraction form was modified after the pilot procedure, to cover the following items: (i) patient sample characteristics, (ii) eligibility criteria, (iii) setting, (iv) procedure and equipment for performing the physical capacity tasks, (v) results of the measurement properties, and (vi) minimal important change. Then, relevant data from the studies included were extracted by one author and checked independently for accuracy by another author.

3.2.7 Assessment of methodological quality

Two authors independently assessed all the articles for methodological quality with the COSMIN 4-point checklist (Mokkink et al., 2010b; Terwee et al., 2012). In this checklist, each measurement property investigated in a study is given a separate rating (excellent, good, fair, or poor) using the "worst score counts method" (Terwee et al., 2012). If consensus could not be reached between the two authors, a third author was consulted to allow consensus to be reached.

3.2.8 Data synthesis

A "best-evidence synthesis" (de Vet et al., 2011) was performed by consensus between all authors. First, the result ratings per measurement property per physical capacity task were determined to be "adequate," "inadequate," or "indeterminate" according to criteria accepted with consensus in an international Delphi study (Table 2) (Prinsen et al., 2016). Second, the level of evidence for the result ratings was assigned as follows (Kroman et al., 2014):

- Strong evidence: consistent result ratings in at least two good-quality articles or at least one excellent-quality article, with a total sample size of eligible articles ≥ 100.

- Moderate evidence: consistent result ratings in at least two fair or one good-quality article, with a total sample size of eligible articles ≥ 50.

- Limited evidence: at least one fair, good, or excellent-quality article, with a total sample size of eligible articles of 25-49.

- Unknown evidence: indeterminate result ratings OR all eligible articles were of poor methodological quality OR total sample size of eligible articles < 25 OR conflicting result ratings.

- Conflicting evidence: conflicting findings.

Multiple studies were only combined in the data synthesis if the studies included a sample with comparable characteristics and if the same measurement property was evaluated for the same physical capacity task. For instance, the data synthesis of tasks that concerned walking was performed separately for samples with patients with back-related diagnoses known to severely affect walking capacity (e.g. lumbar spinal stenosis) and samples without such diagnoses. If an article lacked a priori hypotheses for construct validity and responsiveness, we extracted what the authors had expected in relation to those measurement properties from the description found in the article. We generated our own hypotheses for construct validity and responsiveness based on these descriptions, which were then added to the data synthesis.

Table 2. Criteria for result ratings of measurement properties used in the data synthesis in Study I (Prinsen et al., 2016)

Measurement property	Rating	Criteria for result ratings
Reliability	+	ICC or weighted Kappa ≥ 0.70
	?	ICC or weighted Kappa not reported
	–	Criteria for '+' not met
Measurement error	+	SDC or LoA < MIC*
	?	MIC not defined
	–	Criteria for '+' not met
Hypothesis testing for construct validity	+	75% of the results are in accordance with the hypotheses
	?	No hypotheses defined
	–	Criteria for '+' not met
Criterion validity	+	Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70
	?	Not all information for '+' reported
	–	Criteria for '+' not met
Responsiveness	+	75% of the results are in accordance with the hypotheses OR area under the ROC curve ≥ 0.70
	?	No hypotheses defined
	–	Criteria for '+' not met

*This evidence may come from different studies.

ICC, intraclass correlation coefficient; LoA, limits of agreement; MIC, minimal important change; ROC, receiver operating characteristic; SEM, standard error of measurement; SDC, smallest detectable change; +, adequate rating; –, inadequate rating; ?, indeterminate rating.

3.3 STUDIES II–IV

3.3.1 Protocol and registration

Studies II–IV were based on data from patients participating in a randomized controlled trial (RCT) that compared the effects of a prehabilitation program and conventional preoperative care (Lotzke et al., 2018; Lotzke et al., 2016). The RCT was registered in the ISRCTN registry (www.isrctn.com, registration number: 17115599).

3.3.2 Eligibility criteria

We prospectively included patients aged 18–70 years who were on the waiting lists for lumbar fusion surgery at one university hospital and two private spine clinics. Patients diagnosed with motion-elicited chronic LBP due to DDD in 1–3 segments of the lumbar spine were included. The patients' main surgical procedure was lumbar fusion surgery with the aim of alleviating back pain, but patients could have minor radiating symptoms with or without a simultaneous surgical procedure for lumbar foraminal stenosis, isthmic spondylolisthesis, or disc herniation. We excluded patients who had previously undergone decompression surgery for lumbar spinal stenosis or who had a confirmed

neurological or rheumatic disorder, spinal malignancy, deformities of the thoracolumbar spine such as idiopathic scoliosis, or dominating radiculopathy, or who had a poor understanding of the Swedish language.

3.3.3 Procedure

The patients visited one of three outpatient spine clinics to undergo a clinical examination by an orthopedic surgeon. The orthopedic surgeon made a medical diagnosis based on the clinical and radiological findings, and judged whether lumbar fusion surgery was a treatment option. Patients who agreed to undergo lumbar fusion surgery were placed on a waiting list for surgery. The clinic coordinators continuously informed a physiotherapist in the research group about patients on the waiting list who were potential candidates for inclusion in the study. The physiotherapist then determined whether the patients fulfilled the eligibility criteria. When in doubt about a patient's eligibility, the physiotherapist discussed the patient with the orthopedic surgeons in the project group, to reach consensus.

The physiotherapist contacted eligible patients by telephone and informed them about the study. Patients who were interested in participating met with an independent observer 8–12 weeks before surgery at one of the spine clinics. The independent observer once again provided the patient with information about the study. If the patient agreed to participate, he/she signed an informed consent form provided by the independent observer. Patients were included between April, 2014 and June, 2017.

3.3.4 Data collection

After a patient had agreed to participate, the independent observer guided the participant in performing physical capacity tasks and provided him/her with PROMs and an accelerometer (all outcome measures are described in detail in Section 3.3.5).

Study coordinators contacted patients by telephone to book appointments for follow-up visits 3, 6, 12, and 24 months after surgery. If the study coordinators could not reach a patient, they persevered with telephone calls, voicemail, and e-mail, according to a standardized protocol. If the study coordinators reached a patient but he/she was unable to visit the clinic for follow-up, the study coordinators mailed PROMs and the accelerometer (all outcome measures described in Section 3.3.5). If the study coordinators could not reach a patient, they mailed only the PROMs and not the accelerometer.

Studies II and IV used baseline and 6-month data, while Study III used baseline data. Thus, the follow-up visits 3, 12, and 24 months after surgery were only for the purpose of the RCT (Lotzke et al., 2018; Lotzke et al., 2016).

3.3.5 Intervention

As part of the RCT, the patients were randomly assigned either to participate in a prehabilitation program in preparation for surgery or to receive conventional preoperative care as part of the RCT (Lotzke et al., 2018; Lotzke et al., 2016). In Studies II–IV, the patients were studied irrespective of the preoperative intervention assigned to them.

3.3.6 Outcome measures

Data in Studies II–IV were collected through the use of PROMs, physical capacity tasks, and accelerometers (Table 3).

Table 3. Outcome measures used in Studies II–IV

Outcome measure	Studies II	III	IV
Patient-reported outcome measures			
Oswestry Disability Index 2.0	X	X	X
Visual analog scale for back pain intensity	X	X	X
Visual analog scale for leg pain intensity	X	X	X
Pain Catastrophizing Scale	X	X	X
Tampa Scale for Kinesiophobia	X	X	X
Self-Efficacy for Exercise Scale		X	X
Hospital Anxiety and Depression Scale	X	X	X
Construct-specific global perceived effect scale	X		
Generic global perceived effect scale	X		
Physical capacity tasks			
Five-minute walk	X		
One-minute stair-climbing test	X		X
50-foot walking test	X		X
Timed up-and-go	X		
Measurement of physical activity			
GT3X+ accelerometer		X	X

Demographic data

Gender, age, self-reported height and weight, education level, smoking status, sick-leave status, previous spine surgery, pain duration (back and leg), and comorbidity were collected from the preoperative questionnaire used in Swespine (Fritzell et al., 2018). The type of surgical procedure was obtained from the patients' medical records.

Patient-reported outcome measures

Disability was measured with the Swedish version of the Oswestry Disability Index 2.0 (ODI) (J. C. T. Fairbank et al., 2000). In the ODI, patients rate their perceived disability for ten items concerning pain intensity, personal care, lifting, walking, sitting, standing, sleeping, sex life, social life and traveling. Total scores of the ODI is interpreted as follows: 0–20 represents no disability, 21–40 moderate disability; 41–60 severe disability; 61–80 incapacitating disability; and 81–100 being restricted to bed (J. C. T. Fairbank et al., 1980). The ODI has shown adequate reliability, construct validity, and responsiveness in patients with chronic LBP in a Scandinavian context (Grotle et al., 2004a; Grotle et al., 2003; Lauridsen et al., 2006).

The intensity of back and leg pain over the previous week was measured using 100-mm visual analog scales, with 100 mm meaning maximum pain intensity. There is published support for the validity and reliability of visual analog scales for pain inten-

sity in patients with chronic pain (Carlsson, 1983; Price et al., 1983).

Fear of movement was rated using the Swedish version of the Tampa Scale for Kinesiophobia (TSK). In the TSK, patients rate 17 items assessing subjective thoughts of fear of movement. The Swedish version of the TSK has shown adequate test-retest reliability and internal consistency, and there is support for its face validity, content validity, and construct validity in patients with chronic LBP (Lundberg et al., 2004). Total scores range from 17 to 68, with higher scores indicating a higher level of fear of movement.

Pain catastrophizing was measured using the Swedish version of the Pain Catastrophizing Scale (PCS). In the PCS, patients rate 13 items assessing catastrophizing thoughts about pain. The PCS has shown adequate internal consistency and structural validity for Swedish patients with chronic pain (Kemani et al., 2018). Total scores range from 0 to 52, with 0 meaning no catastrophizing.

Self-efficacy for exercise was measured using the Swedish version of the Self-Efficacy for Exercise Scale (SEES). In the SEES, patients rate their confidence regarding the suggestion that they could exercise three times per week (20 minutes each session) under nine different conditions – for example, "if you would experience pain while you exercised" or "if you would feel tired". The SEES has shown adequate test-retest reliability, internal consistency, and content validity in a Swedish population sample (Rydwik et al., 2014). Total scores range from 0 to 90, with higher scores indicating a higher level of self-efficacy.

Depression was assessed using the Swedish version of the Hospital Anxiety and Depression Scale (HADS). In the HADS,

patients rate seven items related to symptoms of depression. The HADS has shown adequate internal consistency and construct validity in a Swedish population sample (Lisspers et al., 1997).

Perceived postoperative changes in walking ability, stair-climbing ability, and ability to rise from a chair was self-reported with global perceived effect (GPE) scales (which were only used at the 6-month follow-up). The scales comprise the following question and response alternatives (exemplified here for walking ability):

How is your walking ability now compared to how it was before you entered the study?

1. Much better.
2. Better.
3. Somewhat better.
4. Unchanged.
5. Somewhat worse.
6. Worse.
7. Much worse.

Similar GPE scales have been shown to have adequate reliability and validity for patients with chronic LBP (S. J. Kamper et al., 2010; Ward et al., 2015).

Perceived postoperative change in the intensity of back pain were measured with a generic GPE scale (which was only used at the 6-month follow-up) with five response alternatives:

1. Pain-free.
2. Much better.
3. Somewhat better.
4. Unchanged.
5. Worse.

The scale has shown good responsiveness for patients with chronic LBP who undergo lumbar fusion surgery (Hägg et al., 2002).

Physical capacity tasks

In the five-minute walk, participants are asked to walk as fast and as far as possible (without running) for a period of five minutes (Simmonds et al., 1998). The circuit is 15 meters long and octagonal. The distance covered is recorded in meters. The five-minute walk is not to be confused with the 6-minute walk test, which was developed to evaluate pulmonary function in patients with chronic heart failure (Guyatt et al., 1985).

In one-minute stair climbing, the patient is asked to ascend and descend a flight of stairs for one minute, as fast as possible (Smeets et al., 2006). The total number of steps is recorded. The stairs in the current studies had ten steps (each 16 cm high).

In the 50-foot walk, the patient is asked to walk as fast as possible (without running) until he/she gets back to the starting point (Simmonds et al., 1998). The circuit is 15 meters long and octagonal. The time needed to complete the test, in seconds, was recorded.

In timed up-and-go, the patient is asked to rise up from a chair (45 cm high) as fast as possible (without running), walk three meters to a marked line on the floor, turn around, and finally walk back to the chair and sit down (Simmonds et al., 1998). The task is to be performed as quickly as possible. The time needed to complete the test, in seconds, is recorded.

The physical capacity tasks were chosen based on previous research on reliability, validity, and responsiveness, and their clinical usefulness (Andersson et al., 2010; C. E. Lee et al., 2001; Ocarino et al., 2009; Simmonds et al., 1998; Smeets et al., 2006; Teixeira Da Cunha-Filho et al., 2010). At first, we aimed to also include the progressive isoinertial lifting evaluation (Mayer et al., 1988) as a measure of lifting, but we found that the test was too time-consuming to be able to be used in the context of the RCT. We also wanted to include the forward-reach test (Simmonds et al., 1998) as a measure of spinal flexibility, but we found that the test was too difficult to standardize.

Measurement of physical activity

Physical activity was measured with the triaxial accelerometer GT3X+ (ActiGraph, Pensacola, FL, USA). Patients wore the accelerometer for seven consecutive days during waking hours. They were instructed to remove the accelerometer when bathing or swimming. The GT3X+ measures acceleration in three planes, and the raw output is "counts." Based on the number of counts per minute, the raw output can be classified into time spent at different intensities of physical activity by appropriate cut-points in the complementary software, Actilife 6 (ActiGraph). The accelerometer also measures steps per day. The GT3X+ has been shown to have adequate reliability and construct validity for measuring the intensity of physical activity (Kelly et al., 2013; Ozemek et al., 2014) and adequate reliability and criterion validity when measuring the number of steps in healthy adults (Gatti et al., 2016). Moreover, the GT3X+ has shown adequate responsiveness in healthy adults (Swartz et al., 2014).

3.3.7 Statistical analysis

All statistical analysis was performed with SPSS 24.0 for Windows (SPSS Inc., Armonk, NY, USA). An overview of the statistical methods that were used in Studies II–IV is given in Table 4.

Table 4. Overview of the statistical methods used in Studies II–IV

Statistical method	Study II	Study III	Study IV
Descriptive statistics			
Frequency and percentage	X	X	X
Median and interquartile range	X	X	X
Mean and standard deviation	X	X	X
Parametric tests			
Independent-samples t-test	X		X
Non-parametric tests			
Mann-Whitney U-test	X		X
Fisher's exact test	X		X
Responsiveness analysis			
Area under the ROC curve	X		
Spearman correlation	X		
Minimal important change analysis			
Optimal cutoff point of the ROC curve	X		
Physical activity analysis			
Steps per day		X	
Minutes/day of ≥ moderate intensity in 10-minute bouts		X	
Minutes/day of ≥ moderate intensity (non-bouted)			X
Regression analysis			
Univariate linear regression		X	X
Multiple linear regression		X	X

ROC, receiver operating characteristic

Descriptive statistics

Frequencies, means, medians, and their respective variations (percentages, standard deviations, interquartile ranges) were calculated to characterize the participants at baseline (Studies II–IV) and at the 6-month follow-up (Studies II and IV).

Parametric and non-parametric tests

Independent-samples t-test, Mann-Whitney U-test, and Fisher's exact test were used to test for statistically significant baseline differences between patients who were included and excluded in the analyses of responsiveness and minimal important change (Study II) and the prediction analyses (Study IV). The choice of methods depended on the data level and distribution of each variable.

Responsiveness analysis (Study II)

We investigated responsiveness by testing five responsiveness hypotheses (Table 5), as recommended by the developers of COSMIN (de Vet et al., 2011). According to recommendations, an outcome measure is usually considered to have adequate responsiveness

if at least 75% of the hypotheses have been confirmed. In the current study with five hypotheses, we adopted a criterion of at least 80% of the hypotheses being confirmed.

Table 5. Hypotheses for investigating the responsiveness of the physical capacity tasks in Study II

1.	The change scores (differences between baseline and 6-month assessments) of a physical capacity task will be able to distinguish between patients with and without meaningful improvement, as classified by a construct-specific GPE scale (area under the ROC curve ≥ 0.70) (Prinsen et al., 2016; Scott et al., 2015; Ward et al., 2015).[a]
2.	The change scores of a physical capacity task will yield greater misclassifications of improved and unchanged patients in a ROC curve when that classification is based on a generic GPE scale rather than construct-specific GPE scales (Prinsen et al., 2016; Scott et al., 2015; Ward et al., 2015).[a]
3.	The change scores of the four physical capacity tasks will be correlated ≥ 0.50 to each other in the expected direction (Filho et al., 2002; Prinsen et al., 2018; Simmonds et al., 1998; Teixeira Da Cunha-Filho et al., 2010).[b]
4.	The correlations between change scores of physical capacity tasks and the ODI will be at least 0.10 weaker than the correlations between the change scores among the physical capacity tasks themselves (Filho et al., 2002; Gautschi et al., 2016c; C. E. Lee et al., 2001; Prinsen et al., 2018).
5.	The correlations between change scores of a physical capacity task and VAS for intensity of back pain will be at least 0.10 weaker than the correlations between change scores of the physical capacity task and the ODI (Filho et al., 2002; Gautschi et al., 2016c; Prinsen et al., 2018; Simmonds et al., 1998).

[a] For timed up-and-go, hypotheses 1 and 2 were tested separately for the construct-specific GPE scales on walking and chair rise, respectively, since the task includes both of these activities.
[b] The expected direction depends on whether a negative or positive change score of a physical capacity task indicates improvement or deterioration. The correlations between five-minute walk and one-minute stair climbing and also the correlations between 50-foot walk and timed up-and-go were expected to be positive. The other possible correlations among the four physical capacity tasks were expected to be negative.

GPE scale, global perceived effect scale; ROC curve, receiver operating characteristic curve; VAS, visual analog scale

We tested responsiveness hypothesis 1 by calculating the area under the receiver operating characteristic curve for improved and unchanged patients, as classified by the *construct-specific GPE scales* matched for each particular physical capacity task. Improved patients were considered to be those who had scored the response alternatives "much better" or "better" on the construct-specific GPE scales and unchanged patients were those who had scored response alternatives "somewhat better," "unchanged," or "somewhat worse." The area under the receiver operating characteristic curve can vary from 0.5 to 1, and signifies the probability of correctly distinguishing improved patients from unchanged patients, with 1 indicating perfect ability to distinguish improved patients from unchanged patients (de Vet et al., 2011). Hypothesis 1 was accepted if the area under the ROC curve was ≥ 0.70. For timed up-and-go, we tested hypothesis 1 separately for the construct-specific GPE scales on walking and chair rise, respectively, since the task includes both of these activities.

Hypothesis 2 concerned the area under the receiver operating characteristic curve for improved and unchanged patients, as classified by the *generic GPE scale*. Improved patients on the generic GPE scale were considered to be those

who had scored response alternatives "pain-free" or "much better" on the generic GPE scale, and unchanged patients were those who had scored response alternative "somewhat better" or "unchanged." Hypothesis 2 was accepted if the area under the receiver operating characteristic curve for the generic GPE scale was lower than the area under the area under the receiver operating characteristic curve for the construct-specific GPE scales. For timed up-and-go, we tested hypothesis 2 separately for the construct-specific GPE scales on walking and chair rise.

Hypotheses 3–5 were investigated with Spearman's rho, as the change scores of the physical capacity task were not normally distributed (de Winter et al., 2016).

Minimal important change analysis (Study II)

We determined the minimal important change (MIC) for improvement (hereafter referred to only as MIC) with the optimal cutoff point of the ROC curve based on the classification of improved and unchanged patients according to construct-specific GPE scales. We matched the GPE scales for each specific physical capacity task. MICs for timed up-and-go were calculated separately for the construct-specific GPE scales for walking and chair rise. The optimal cutoff point of the ROC curve represents the change score of each physical capacity task that yields the smallest number of misclassifications between improved and unchanged patients (de Vet et al., 2006). We calculated absolute MICs based on the ROC curve plotted with the absolute change scores of each physical capacity task. We calculated relative MICs based on the ROC curve plotted with the percentage of change from baseline for each physical capacity task. We did not calculate MIC for deterioration, since only two patients reported deterioration on the construct-specific GPE scales.

Physical activity analysis (Studies III and IV)

We processed accelerometer data on a minute-by-minute basis in the software Actilife, to generate three variables:

1. Steps per day.

2. Minutes per day of at least moderate-intensity physical activity accumulated in at least 10-minute bouts. A 10-minute bout was defined as a 10-minute period with an interruption of no more than two minutes below the threshold of 2,020 counts (Troiano et al., 2008).

3. Minutes per day of at least moderate-intensity physical activity regardless of whether or not the physical activity was performed in 10-minute bouts (non-bouted) (Troiano et al., 2008).

In Study III, the variable steps per day was used as a measure of total physical activity regardless of the intensity at which it was performed. Minutes per day of at least moderate-intensity physical activity (in 10-minute bouts) was used as a measure of the intensity of physical activity. The rationale for the 10-minute bout criterion was to enable comparisons with the WHO recommendations on physical activity (World Health Organization, 2009b). To further enable comparisons with the recommendations, the 10-minute bout variable was multiplied by seven to yield the number of minutes per week.

In Study IV, the variable minutes per day of at least moderate-intensity physical activity (non-bouted) was used. The rationale for using the non-bouted variable instead of the 10-minute bout variable was that the results of Study III showed that the 10-minute bout variable was not appropriate for regression analysis (see Section 4.4.2).

Patients had to have ≥ 4 days of ≥ 10 hours wear time per day of the accelerometer to be included in further analysis (Trost et al., 2005). Wear time was defined according to Choi et al. (2011).

Adherence to physical activity recommendations (Study III)

We calculated the proportions of patients who reached the WHO physical activity recommendations (≥ 150 minutes per week of at least moderate-intensity physical activity performed in bouts of at least 10 minutes) by dividing the number of patients who fulfilled the recommendations by the total number of patients with sufficient wear time. In the same manner, we calculated the proportions of patients who achieved ≥ 7,500 steps per day (physically active lifestyle), 5,000–7,499 steps per day (low active lifestyle), or < 5,000 steps per day (sedentary lifestyle) (Tudor-Locke et al., 2013).

Regression analysis investigating associations between fear-avoidance variables and health outcomes (Studies III and IV)

In Study 3, we investigated the associations between fear-avoidance variables and the preoperative level of physical activity in two multiple linear regression models: one model with steps per day as the dependent variable and one with minutes per week of at least moderate-intensity physical activity (in 10-minute bouts) as the dependent variable (Table 6). The selection of potential independent variables in the models was based on the modified fear-avoidance model of Lotzke et al. (2016). The maximum number of potential independent variables was set to ten based on the sample size of 118 (power level = 0.8, alpha level = 0.05, effect size = 0.15 ("moderate")) (Cohen, 1988).

Table 6. Overview of the potential independent variables and the dependent variables in the regression models in Study III

Potential independent variables
Baseline disability (ODI)
Baseline back pain intensity (VAS)
Baseline leg pain intensity (VAS)
Baseline fear of movement (TSK)
Baseline pain catastrophizing (PCS)
Baseline self-efficacy for exercise (SEES)
Baseline depression (HADS)
Confounder: Baseline age
Confounder: Baseline BMI
Confounder: Gender
Dependent variables
Model 1: Steps per day (GT3X+)
Model 2: Minutes per week of at least moderate-intensity physical activity (in 10-minute bouts) (GT3X+)

HADS, Hospital Anxiety and Depression Scale; ODI, Oswestry Disability Index 2.0; PCS, Pain Catastrophizing Scale; SEES, Self-Efficacy for Exercise Scale; TSK, Tampa Scale for Kinesiophobia; VAS, visual analog scale.

In Study IV, we investigated the associations between potential predictors and change scores (the difference between baseline and 6-month assessments) of physical activity level and disability in two separate multiple linear regression models: one model with minutes per day of at least moderate-intensity physical activity (non-bouted) as the dependent variable and one with disability as the dependent variable (Table 7). We based the selection of the predictor variables on the modified version of the fear-avoidance model as described by Lotzke et al. (2016) (Figure 2). Other variables were included, based on the empirical evidence presented in other studies investigating predictors for the outcome of lumbar spine surgery (Abbott et al., 2011; Carreon et al., 2009; Gunzburg et al., 2003; LaCaille et al., 2005; Mannion et al., 2006; Soriano et al., 2010; Trief et al., 2000; Trief et al., 2006).

Table 7. Overview of the potential predictors and the dependent variables in the prediction models in Study IV

Potential predictors
Baseline value of the dependent variable
Baseline back pain intensity (VAS)
Baseline leg pain intensity (VAS)
Baseline pain catastrophizing (PCS)
Baseline fear of movement (TSK)
Baseline self-efficacy for exercise (SEES)
Baseline depression (HADS)
Baseline walking capacity (five-minute walk)
Baseline walking capacity (50-foot walk)
Baseline sick leave [yes/no]
Baseline duration of back pain [≤ 2 years/> 2 years]
Baseline age
Baseline BMI
Gender
Dependent variables
Physical activity model: Postoperative change from baseline in minutes per day of at least moderate-intensity physical activity (non-bouted) (GT3X+)
Disability model: Postoperative change from baseline in disability level (ODI)

HADS, Hospital Anxiety and Depression Scale; ODI, Oswestry Disability Index 2.0; PCS, Pain Catastrophizing Scale; SEES, Self-Efficacy for Exercise Scale; TSK, Tampa Scale for Kinesiophobia; VAS, visual analog scale.

In both Study II and Study IV, we used a purposeful selection method with three steps to select which of the independent variables would be retained in the final regression models (Bursac et al., 2008; Hosmer et al., 1999):

1. Only potential independent variables associated with the dependent variable at a p-value ≤ 0.25 in univariate regression analysis were brought forward to the second step of the analysis.

2. The remaining independent variables were included in a backward multiple regression analysis along with the baseline equivalent of the dependent variable. Independent variables with a p-value > 0.15 were removed in the backward multiple regression analysis if the beta coefficients of the remaining independent variables did not change more than 15%.

3. The independent variables that were excluded in the first step were added back to the multiple regression model one by one, and were only kept if they had a p-value ≤ 0.15.

The independent variables in the final models were controlled for multicollinearity with correlation analysis (Pearson's r or Spearman rho depending on data level and distribution). The standardized residuals from the regression models were checked for normality and heteroscedasticity with normality and heteroscedasticity plots (Fahrmeir et al., 2013).

In Study III, the findings of any remaining confounders (age, gender, or BMI) in the final models were not interpreted, as they were only added to adjust the model. In Study IV, age, gender, and BMI were considered to be potential predictors and not confounders.

SUMMARY OF RESULTS

4.1 STUDY I

The aim of Study I was to systematically review the level of evidence of reliability, validity, and responsiveness of physical capacity tasks that are designed to assess functioning in patients with LBP.

4.1.1 Study selection

The electronic search and hand search resulted in 7,900 articles after the removal of duplicates (Figure 4). Twenty-five of these were included (Andersson et al., 2010; Armstrong et al., 2005; Campbell et al., 2006; Conway et al., 2011; Deen et al., 2000; Gautschi et al., 2016a, 2016b; Gautschi et al., 2017; Kahraman et al., 2016; C. E. Lee et al., 2001; Magnussen et al., 2004; Ocarino et al., 2009; Odebiyi et al., 2007; Pratt et al., 2002; Rainville et al., 2012; Simmonds et al., 1998; Smeets et al., 2006; Soer et al., 2006; Staartjes et al., 2018; Strand et al., 2011; Strand et al., 2002; Taylor et al., 2001; Teixeira Da Cunha-Filho et al., 2010; Tomkins et al., 2009; Whitehurst et al., 2001). One article comprised patients with a pain duration of ≥ 6 weeks (subacute LBP) (Magnussen et al., 2004), while the remaining 24 articles comprised patients with a pain duration of ≥ 12 weeks (chronic LBP). Eleven of the articles comprised patients with chronic LBP who had back-related diagnoses known to severely affect walking capacity. Of these, one article included patients with lumbar spondylolisthesis and post-laminectomy, six articles included patients with lumbar spinal stenosis, three articles included patients with lumbar disc herniation and lumbar spinal stenosis, and one article included patients with lumbar disc herniation, lumbar spinal stenosis, and spondylolisthesis.

| 12,301 articles identified through the first database search [Inception – May 11, 2017] | 2,302 articles identified through the second database search [Jan 1, 2016 – Aug 29, 2018] | 1,557 articles identified through hand search of 362 reference lists |

7,900 articles after duplicates removed

| Abstracts of 7,900 articles screened | → | 7,440 articles excluded |

| 460 full-text articles assessed for eligibility | → | 435 full-text articles excluded: |

1. Original data not included: 201
2. Wrong condition or not fulfilling inclusion criteria for age: 47
3. Tests not adequately described or original reference for test procedure not obtained: 14
4. Wrong construct: 65
5. Tests not evaluated by an observer: 15
6. Insufficient standardization/expensive equipment: 18
7. Aim was not to evaluate measurement properties: 69
8. Results only presented for composite score of a test battery: 5

Full-text articles not found: 1

25 articles included in the data extraction (of which 5 were from hand search)

25 articles included in the methodological quality assessment

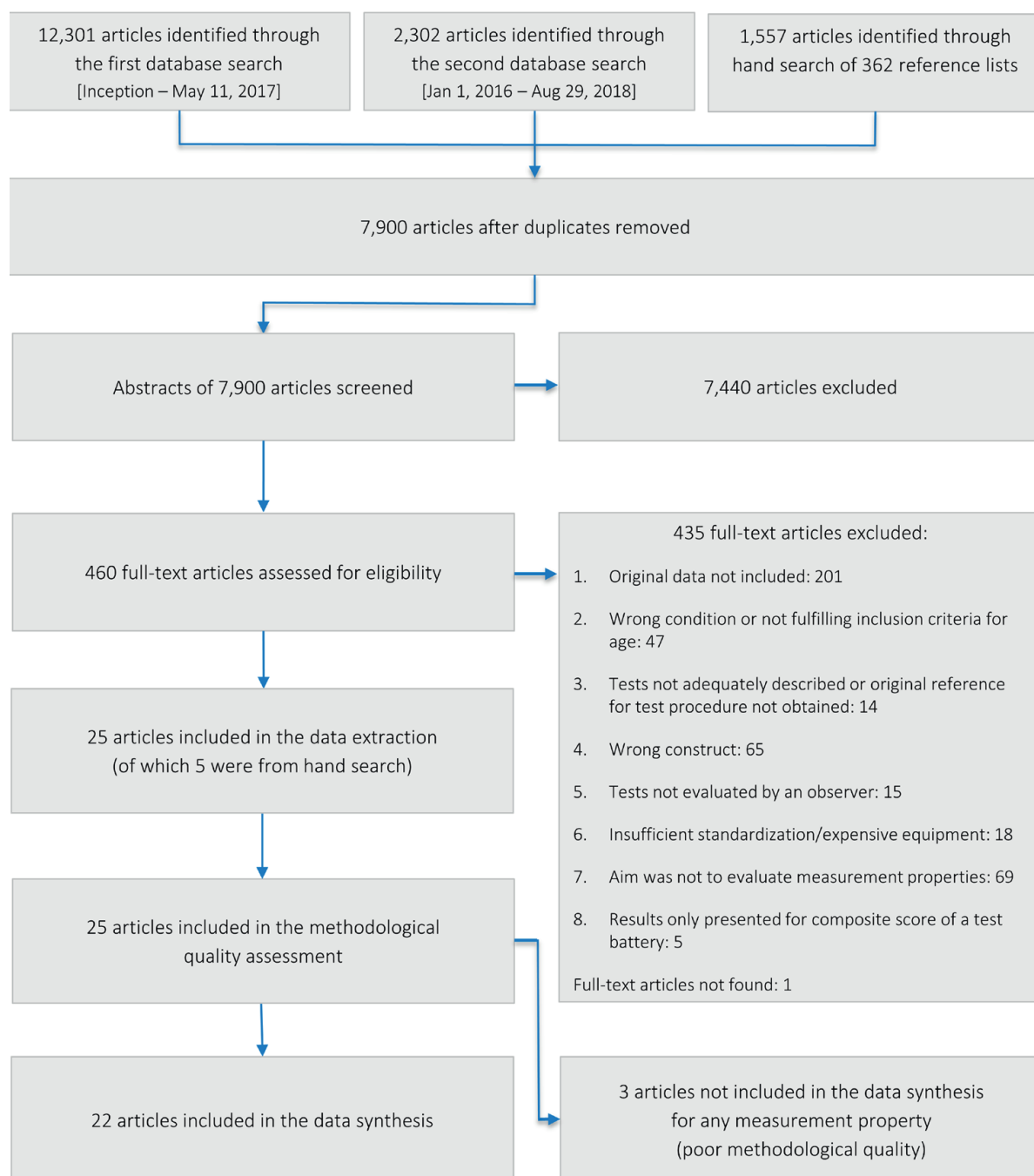| 22 articles included in the data synthesis | → | 3 articles not included in the data synthesis for any measurement property (poor methodological quality) |

Figure 4. Flow chart of article inclusion in Study I.

The 25 articles included covered 18 physical capacity tasks (Table 8). The physical capacity tasks involved the following activities: walking (10 tasks), stair climbing (1 task), lifting (3 tasks), rising from a chair (2 tasks), a combination of walking/rising from a chair (1 task), and a combination of walking and carrying (1 task).

Table 8. Brief descriptions of the physical capacity tasks that were included in Study I.

Physical capacity task	Quantification measure	Equipment needed
One-minute stair climbing	Number of stairs climbed in 1 minute	A flight of stairs with handrails, stopwatch
30-second chair stand test	Number of repetitions (sitting → standing) performed in 30 seconds	Chair, stopwatch
Five-repetition sit-to-stand	Seconds to complete five repetitions of sitting to standing	Chair, stopwatch
50-foot walk	Seconds to complete a 50-foot course at maximum speed	Measuring tape, stopwatch, markers for indicating track endpoints
50-foot walk, preferred speed	Seconds to complete a 50-foot course at preferred speed	Measuring tape, stopwatch, markers for track endpoints
Five-minute walk	Meters walked in 5 minutes	Measuring tape, stopwatch, markers for indicating track endpoints
Lift test	Ordinal scale of the number of repetitions lifting box with sandbag from floor to table and back	Table, 1.35 kg box with 5-kg sandbag
Lift test, modified	Number of repetitions lifting a box from floor to table and back in 1 minute	Table, 1.35 kg box with 5-kg sandbag, 4-kg for women
Motorized treadmill test	Total walking time and distance walked at preferred speed [non-modifiable during the test] at the moment walking-related symptoms make the participant stop	Treadmill, stopwatch
Progressive isoinertial lifting evaluation	Weight in kg of the box during the last completed cycle (Strand et al., 2011)/Number of completed lifting cycles (Andersson et al., 2010; Smeets et al., 2006)	Standardized box and an assortment of weights
Self-paced walking test	Total walking time, speed, and distance walked at the moment walking-related symptoms make the participant stop	Distance instrument, stopwatch
Shuttle walking test	Number of meters walked until the participants fail to complete a predefined "shuttle" in the time allocated	Standardized audio tape, measuring tape, markers for indicating track endpoints
Timed up-and-go	Seconds to complete rising up from a chair, walking 3 meters, turning around, walking back to the chair, and sitting down.	Chair, stopwatch
Treadmill examination, 1.2 miles per hour	Total time walked at 1.2 miles per hour on a treadmill at the moment walking-related symptoms make the participant stop [time limit: 15 minutes]	Treadmill, stopwatch

(cont.)

	(Table 8 cont.)	
Treadmill examination, preferred speed	Total time walked at preferred speed on a treadmill at the moment walking-related symptoms make the participant stop [time limit: 15 minutes]	Treadmill, stopwatch
Treadmill protocol	Total distance, time, and average speed walked on a treadmill at preferred speed [modifiable during the test] the moment symptoms of lumbar spinal stenosis or other reasons make the participant stop [time limit: 30 minutes]	Treadmill, stopwatch
Treadmill walking test	Distance walked on a treadmill at 53.6 meters/minute at the moment that pain or fatigue make the participant stop, or when 70% of the predicted maximum heart rate (70 [age − 220]) is reached	Treadmill, stopwatch, heart rate monitor
Weight-carrying test	Time needed to walk 20 meters as quickly as possible while carrying dumb-bells weighing equivalent to 10% of the person's weight	Stopwatch, an assortment of dumb-bells

4.1.2 Assessment of methodological quality

The majority of the articles were scored fair for methodological quality. Three studies of reliability in one of the articles (Strand et al., 2011) and nine studies of measurement error in three of the articles (Andersson et al., 2010; Gautschi et al., 2017; Strand et al., 2011) were rated as poor and therefore excluded from the best evidence synthesis. The poor scores were due to the fact that patients received treatment between the first and second administration of the tasks. Five studies of construct validity in two of the articles (Ocarino et al., 2009; Odebiyi et al., 2007) were rated as poor and therefore excluded from the best evidence synthesis. The poor scores were due to an absence of a priori validity hypotheses (such as the hypothesized correlation between two outcome measures required for adequate validity).

4.1.3 Data synthesis

The articles investigated five measurement properties: reliability, measurement error, construct validity (hypothesis testing), criterion validity, and responsiveness. The physical capacity tasks that had adequate re-

sults for more than one measurement property are summarized below.

Five-repetition sit-to-stand was the only physical capacity task that had adequate ratings for more than two measurement properties: test-retest reliability (strong evidence), construct validity (moderate evidence), and responsiveness (limited evidence) (Table 9). Fifty-foot walk, five-minute walk, progressive isoinertial lifting evaluation, and timed up-and-go showed moderate to strong evidence for adequate test-retest reliability and construct validity. The above-mentioned tasks, however, also showed moderate to strong evidence for inadequate measurement error. One-minute stair climbing and shuttle walking test displayed adequate responsiveness (limited evidence) as well as adequate test-retest reliability (moderate evidence for one-minute stair climbing and limited evidence for shuttle walking test).

The level of evidence for walking tasks that were investigated for patients with diagnoses known to severely affect walking capacity is marked in italics in Table 9. Of these tasks, timed up-and-go and shuttle walking

test were the only ones that showed adequate ratings for more than one measurement property. The level of evidence for those ratings was, however, limited.

Table 9. Summary of the level of evidence per measurement property of the physical capacity tasks included in Study I

Physical capacity task	Test-retest reliability	Inter-rater reliability	Intra-rater reliability	Measurement error	Construct validity (hypothesis testing)	Criterion validity	Responsiveness
One-minute stair climbing	Moderate (+)	0	0	Moderate (−)	0	0	Limited (+)
30-second chair stand	0	0	Limited (+)	0	Limited (+)	0	0
Five-repetition sit-to-stand	Strong (+)	Unknown	0	Strong (−)	Moderate (+)	0	Limited (+)
50-foot walk	Strong (+)	Unknown	0	Strong (−)	Moderate (+)	0	Moderate (−)
50-foot walk, preferred speed	Conflicting	Unknown	0	Unknown	Conflicting	0	0
Five-minute walk	Strong (+)	0	0	Moderate (−)	Moderate (+)	0	Limited (−)
Lift test	Limited (−)	Limited (+)	0	0	0	0	Limited (−)
Lift test, modified	Unknown	0	0	Unknown	Moderate (+)	0	Moderate (−)
Motorized treadmill test	0	0	0	0	*Limited (+)**	0	Limited (−)
Progressive isoinertial lifting evaluation	Moderate (+)	0	0	Moderate (−)	Moderate (+)	0	Moderate (−)
Self-paced walking test	*Limited (+)**	0	0	0	*Conflicting**	0	*Limited (−)**
Shuttle walking test	Limited (+), *Limited (+)**	0	0	Unknown	0	0	Limited (+), *Limited (+)**
Timed up-and-go	Moderate (+)	Unknown	0	Unknown	Moderate (+), *Limited (+)**	0	*Limited (+)**
Treadmill examination, 1.2 miles per hour	*Limited (+)**	0	0	0	0	0	0
Treadmill examination, preferred speed	*Limited (+)**	0	0	0	0	0	0
Treadmill protocol	0	0	0	0	0	*Limited (+)**	0
Treadmill walking test	0	0	0	0	*Limited (+)**	0	0
Weight-carrying test	0	0	0	0	*Limited (+)**	0	0

*The level of evidence in *italics* primarily concerns patients with back-related diagnoses that are known to severely affect walking capacity (e.g. lumbar spinal stenosis and spondylolisthesis);

+, adequate result rating; −, inadequate result rating; 0, no information

4.2 PATIENTS IN STUDIES II–IV

Studies II–IV included 63 women and 55 men (Figure 5) with a mean age of 45.7 years (Table 10). Of these, 108 underwent lumbar fusion surgery, while seven declined surgery and three underwent other lumbar spine surgery. The proportion of men and women and the patients' mean age were similar to those of patients with chronic LBP due to DDD in Swespine (Fritzell et al., 2018). The patients' baseline levels of disability and intensity of back pain were lower compared to those in the Swedish Spine Registry (Fritzell et al., 2018) (Table 11).
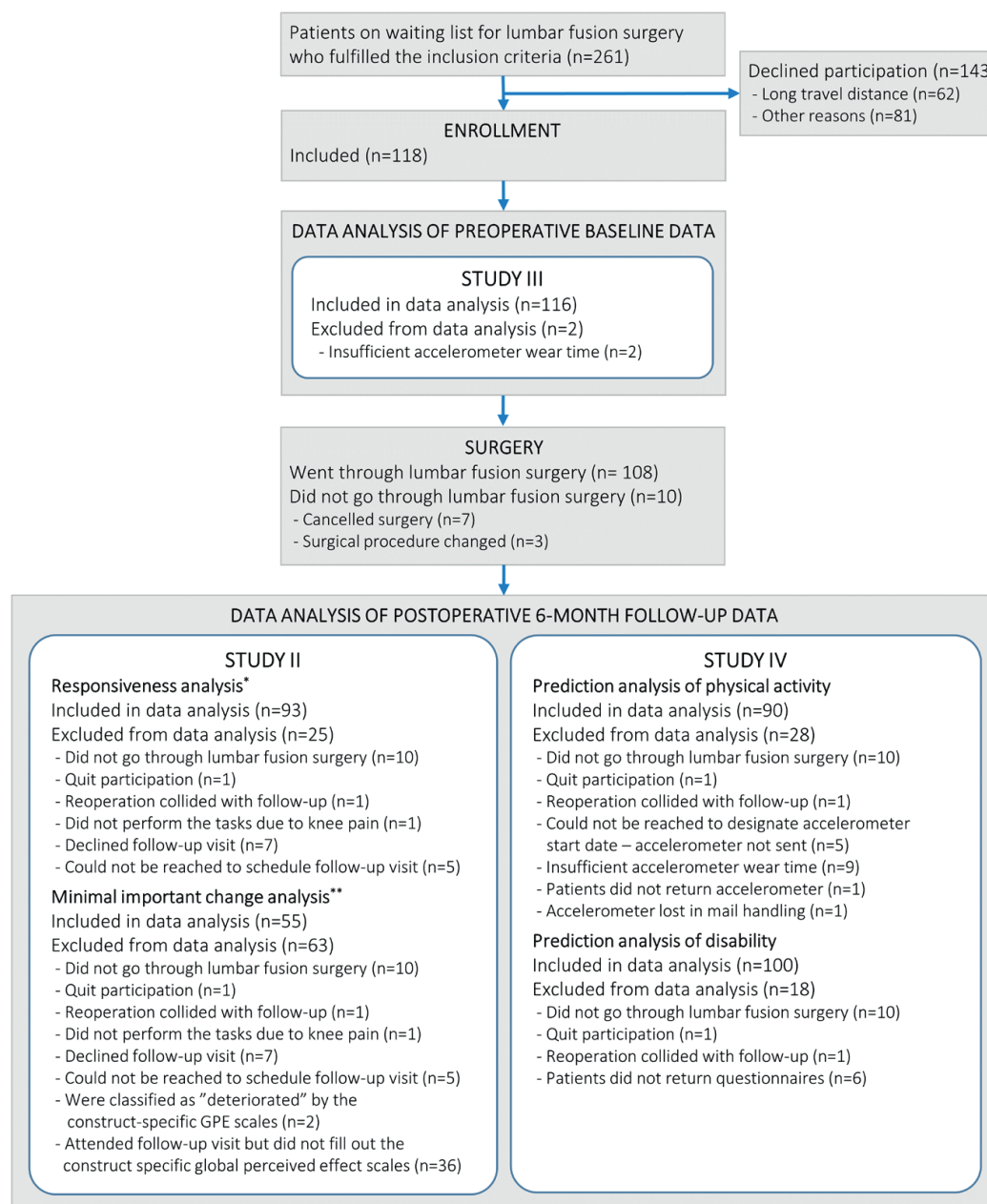

Figure 5. Flow chart of patient inclusion and data analysis in Studies II–IV.

*The number of patients concerns the responsiveness analysis for hypotheses 2–5. In the data analysis for responsiveness hypothesis 1, only 57 patients were included due to missing data on the construct-specific global perceived effect scales. Moreover, since one patient had missing baseline data for 5-minute walk, the number of patients in the responsiveness analysis for all hypotheses was one less for this task than for the others.
**For 5-minute walk, 54 patients were included in the minimal important change analysis since one patient had missing baseline data for that physical capacity task.

Table 10. Sociodemographics of the patients included in Studies II–IV

Variable	All patients (n = 118)
Age, mean (SD)	45.7 (8.3)
Gender, n (%)	
Women	63 (53.4%)
Men	55 (46.6%)
Body mass index, mean (SD)	26.3 (3.7)
Education level, n (%)	
Elementary school	7 (6.0%)
High school	51 (43.2%)
University or college	42 (35.6%)
Vocational education	17 (14.4%)
Missing information	1 (0.8%)
Work status, n (%)	
Working	73 (61.9%)
Partial sick leave/absence	15 (12.7%)
Fulltime absence	22 (18.7%)
Unemployed	5 (4.2%)
Missing information	3 (2.5%)
Back pain duration, n (%)	
3–12 months	9 (7.6%)
> 1 year to ≤ 2 years	20 (17.0%)
> 2 years	87 (73.7%)
Missing information	2 (1.7%)
Previous lumbar spine surgery, n (%)	
Yes	11 (9.3%)
No	107 (90.7%)
Current surgical procedure, n (%)	
Instrumented posterior fusion	102 (86.5%)
Instrumented combined posterior and interbody fusion	5 (4.2%)
Instrumented anterior interbody fusion	1 (0.8%)
Did not go through lumbar fusion surgery	10 (8.5%)

SD, standard deviation.

Table 11. Baseline characteristics of the patients included in Studies II–IV

Variable	All patients (n = 118)	Reference values
Patient-reported outcome measures		
Disability, ODI, mean (SD)	37.8 (12.4)	Swespine: 43.0 (Fritzell et al., 2018)
Back pain intensity, VAS, mean (SD)	61.1 (19.4)	Swespine: 69.0 (Fritzell et al., 2018)
Leg pain intensity, VAS, mean (SD)	35.4 (29.7)	Swespine: 42.0 (Fritzell et al., 2018)
Pain catastrophizing, PCS, mean (SD)	22.8 (8.1)	$\geq$ 20: threshold for pain catastrophizing (Sullivan et al., 2006)
Fear of movement, TSK, mean (SD)	38.1 (8.4)	$\geq$ 37: threshold for kinesiophobia (Lundberg et al., 2004)
Self-efficacy for exercise, SEES, mean (SD)	61.2 (20.5)	–
Depressed mood, HADS, mean (SD)	5.4 (3.6)	0–7: normal level of depression (Zigmond et al., 1983)
Physical capacity tasks		
Five-minute walk, mean (SD)	418.6 (81.8)	Pain-free population, USA: 518 meters (Simmonds et al., 1998)
One-minute stair climbing, mean (SD)	104.1 (24.6)	–
50-foot walk, mean (SD)	9.3 (2.8)	Pain-free population, USA: 8.4 seconds (Simmonds et al., 1998)
Timed up-and-go, mean (SD)	7.9 (3.0)	Pain-free population, USA: 5.2 seconds (Simmonds et al., 1998)
Physical activity		
Minutes per week of at least moderate-intensity physical activity [non-bouted], GT3X+, mean (SD)	197.6 (141.3)	Swedish norm values for individuals aged 40–59 years: 233.1 minutes* (Hagströmer et al., 2010)
Minutes per week of at least moderate-intensity physical activity [in 10-minute bouts], GT3X+, mean (SD)	81.7 (116.9)	WHO recommendations: $\geq$ 150 minutes per week (World Health Organization, 2009b)
Steps per day, mean (SD)	7,493 (2,645)	< 5,000: 'sedentary lifestyle'; 5,000–7,499: 'low active lifestyle'; $\geq$ 7,500: 'physically active lifestyle' (Tudor-Locke et al., 2013)

*The original article presented mean values for men and women separately (Hagströmer et al., 2010). For the purpose of this table, the separate means for men and women were recalculated to give a single mean.

HADS, Hospital Anxiety and Depression Scale; IQR, interquartile range; ODI, Oswestry Disability Index 2.0; PCS, Pain Catastrophizing Scale; SD, standard deviation; SEES, Self-Efficacy for Exercise Scale; Swespine, the Swedish National Quality Registry for Spine Surgery (preoperative values for patients with degenerative disc disease who underwent lumbar fusion in 2017 are presented); TUG, timed up-and-go; TSK, Tampa Scale for Kinesiophobia; VAS, visual analog scale.

4.3 STUDY II

The aim of Study II was to investigate the responsiveness and MIC of four physical capacity tasks used to assess functioning in

patients with chronic LBP due to DDD who undergo lumbar fusion surgery.

4.3.1 Responsiveness

One-minute stair climbing, 50-foot walk, and timed up-and-go showed adequate responsiveness (80% of the hypotheses confirmed for one-minute stair climbing and 50-foot walk, and 100% for timed up-and-go), whereas five-minute walk did not (40% of the hypotheses confirmed) (Table 12).

Table 12. Summary of the results of responsiveness hypothesis testing of the four physical capacity tasks in Study II

Physical capacity tasks	Construct-specific GPE used in hypotheses 1 & 2	Results of hypothesis testing					Number of hypotheses confirmed
		1	2	3	4	5	
Five-minute walk	GPE$_{walking}$	−	−	+	+	−	2 of 5 = 40%
One-minute stair climbing	GPE$_{stair\ climbing}$	+	+	+	+	−	4 of 5 = 80%
50-foot walk	GPE$_{walking}$	+	+	+	+	−	4 of 5 = 80%
Timed up-and-go*	GPE$_{walking}$	+	+	+	+	+	5 of 5 = 100%
	GPE$_{chair\ rise}$	+	+				

*For timed up-and-go, hypotheses 1 and 2 were tested separately for the construct-specific GPE scales on walking and chair rise, respectively, since the task includes both of these activities.

−, rejected hypothesis; +, confirmed hypothesis; GPE, global perceived effect scale.

4.3.2 Minimal important change

Absolute MICs were 45.5 meters (five-minute walk), 20 steps (one-minute stair climbing), −0.6 seconds (50-foot walk), and −1.3 seconds (timed up-and-go). Absolute and relative MICs along with their associated sensitivity and specificity are presented in Table 13.

Table 13. Absolute and relative values for minimal important change (MIC) and their associated sensitivity and specificity for the four physical capacity tasks in Study II

Physical capacity tasks	5-minute walk	1-minute stair climbing	50-foot walk	Timed up-and-go (GPE$_{walking}$)	Timed up-and-go (GPE$_{chair\ rise}$)
MIC$_{absolute}$ (95% CI)	45.5 m (8.5 to 62.0)	20.0 steps (10.5 to 48.0)	−0.6 s (−0.7 to −0.2)	−1.3 s (−2.4 to −0.5)	−1.3 s (−2.4 to −0.3)
Sensitivity	0.63	0.63	0.73	0.67	0.73
Specificity	0.75	0.76	0.83	0.71	0.79
MIC$_{relative}$ (95% CI)	9.0% (4.5 to 11.8)	12.5% (7.2 to 48.4)	−6.1% (−7.1 to −3.4)	−17.3% (−29.4 to −10.2)	−17.6% (−20.7 to −10.2)
Sensitivity	0.69	0.70	0.76	0.73	0.79
Specificity	0.75	0.72	0.88	0.71	0.79

GPE, global perceived effect scale

4.4 STUDY III

The aim of Study III was to investigate the preoperative level of physical activity in patients with chronic LBP due to DDD scheduled for lumbar fusion surgery, and furthermore to investigate whether fear-avoidance variables were associated with this level.

4.4.1 Preoperative level of physical activity

Twenty patients (17%) fulfilled the WHO recommendations on physical activity for health. Thirty-two patients (28%) spent zero minutes per week on moderate-intensity physical activity (in 10-minute bouts) and 64 patients (55%) spent between 1 and 149 minutes per week on moderate-intensity physical activity (in 10-minute bouts) (Figure 6).
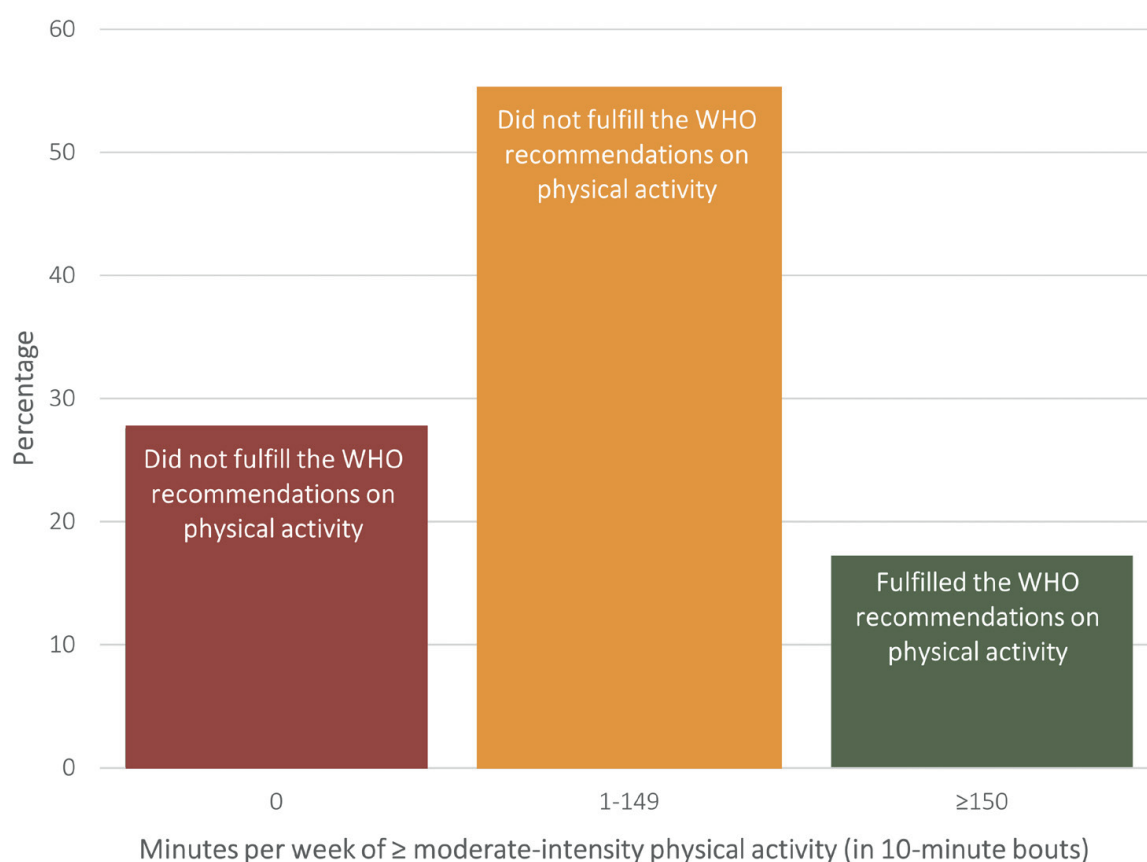

Figure 6. Histogram of the preoperative level of physical activity (measured as minutes per week of least moderate- intensity physical activity (in 10-minute bouts)) in Study III.

Nineteen patients (16%) walked less than 5,000 steps per day (sedentary lifestyle), 44 patients (38%) between 5,000 and 7,499 steps per day (low active lifestyle), and 53 patients (46%) walked ≥ 7,500 steps (physically active lifestyle) (Figure 7).
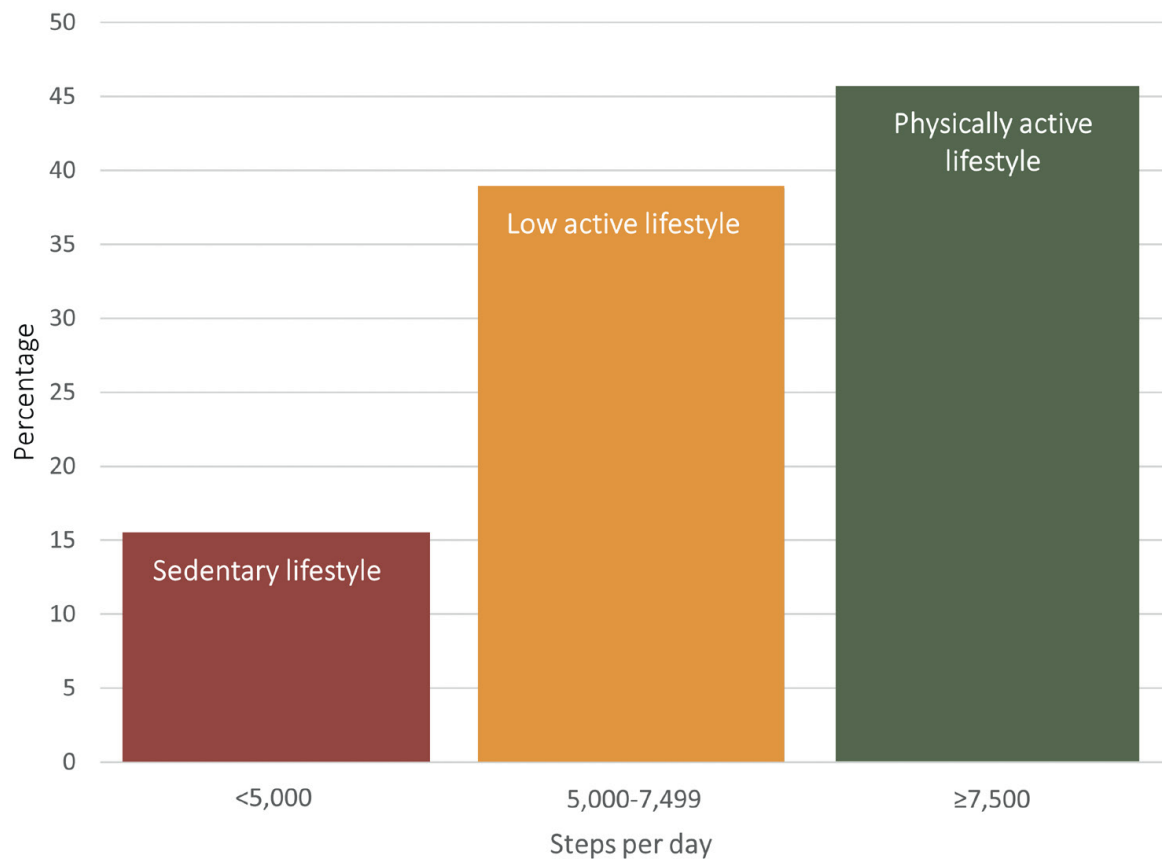
Figure 7. The preoperative number of steps per day according to the classification of steps per day by Tudor-Locke et al. (2013) (Study III).

4.4.2 Associations between fear-avoidance variables and the preoperative level of physical activity

The standardized residuals in the multiple linear regression analysis of steps per day and minutes per week of at least moderate-intensity physical activity (in 10-minute bouts) were not normally distributed, and the variables were therefore transformed into their natural logarithms. The standardized residuals of minutes per week of at least moderate-intensity physical activity (in 10-minute bouts) were still not normally distributed after the transformation, and the variable was therefore not investigated further.

The final regression model for steps per day (log-transformed) as the dependent variable

contained back pain, fear of movement, disability, and BMI (Table 14) (R^2 = 17%). Of these, fear of movement, disability, and BMI were found to be significantly negatively associated with the dependent log-transformed variable steps per day ($p < 0.05$). *At the* group level, a 10-point lower level of fear of movement (TSK) was associated with an 8.6% greater number of steps per day, as was a 10-point lower level of disability (percentages given by back-transforming the unstandardized beta coefficients). The results for BMI were not interpreted, as the variable was only added to adjust the model.

Table 14. Results of multiple linear regression analysis with the dependent variable steps per day (log-transformed) in Study III ($R^2 = 17\%$)

Variables	Unstandardized beta	p-value	95% CI for beta	
			Lower	Upper
Constant	9.930	0.000	9.390	10.470
ODI	−0.009	0.006	−0.015	−0.003
TSK	−0.009	0.034	−0.017	−0.001
BMI	−0.021	0.020	−0.038	−0.003
VAS$_{Back}$	0.002	0.293	−0.002	0.006

BMI, body mass index; CI, confidence interval; ODI, Oswestry Disability Index; VAS$_{Back}$, visual analog scale for back pain intensity; TSK, Tampa Scale for Kinesiophobia.

4.5 STUDY IV

The aim of Study IV was to investigate the predictive value of preoperative fear-avoidance variables, walking capacity, and traditional predictor variables for prediction of postoperative changes in physical activity level and disability six months after lumbar fusion surgery in patients with chronic LBP due to DDD.

4.5.1 Prediction of physical activity

The preoperative levels of physical activity and self-efficacy for exercise were significant predictors of the change in the level of physical activity from baseline to the 6-month follow-up ($R^2 = 25.1\%$) (Table 15). These results indicate that patients with a low preoperative level of physical activity were more likely to increase their level of postoperative physical activity, relative to those with a higher preoperative level of activity. This relationship is illustrated in Figure 8. The model also demonstrated that patients with high preoperative self-efficacy for exercise were more likely to increase their postoperative physical activity level.

Table 15. Overview of the prediction models for physical activity and disability in Study IV

Model	Predictor	Unstandard-ized beta	p-value	95% CI for beta		R^2
				Lower	Upper	
Model 1: Physical activity	Constant	−0.081	0.987	−9.931	9.770	–
	Preoperative physical activity level	−0.349	< 0.001	−0.482	−0.216	Partial R^2 = 20.4%
	Preoperative self-efficacy for exercise	0.176	0.021	0.027	0.325	Partial R^2 = 4.7%[a]
						Total R^2 = 25.1%

(cont.)

		(Table 15 cont.)				
Model 2: Disability	Constant	−7.150	0.287	−20.416	6.116	−
	Preoperative disability	−0.790	0.000	−1.026	−0.553	Partial R^2 = 27.8%
	Preoperative	0.152	0.024	0.020	0.284	Partial R^2 = 3.1%[b]
	Preoperative pain catastrophizing	0.383	0.030	0.038	0.728	Partial R^2 = 3.3%[c]
						Total R^2 = 34.2%

[a]Refers to the explained variance above the explained variance of the physical activity model when only the preoperative level of physical activity was included.
[b]Refers to the explained variance above the explained variance of the disability model when only the preoperative level of disability was included.
[c]Refers to the explained variance above the explained variance of the disability model when only the preoperative level of disability and self-efficacy for exercise were included.

CI, confidence interval; ODI, Oswestry Disability Index; R^2, explained variance; SD, standard deviation.
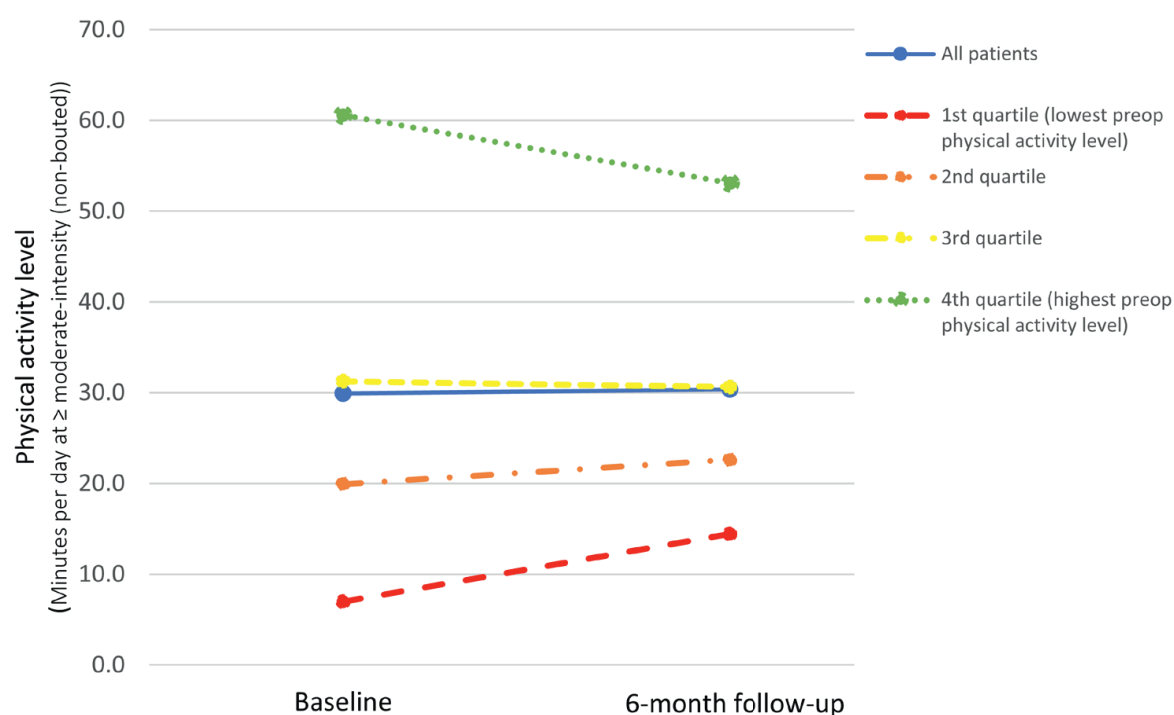

Figure 8. Scores at baseline and at 6-month follow-up for physical activity level (minutes per day of at least moderate-intensity physical activity (non-bouted)), for all patients and for quartiles based on the patient's preoperative physical activity level (Study IV).
Reproduced from World Neurosurgery, CC permission.

4.5.2 Prediction of disability

Preoperative disability, self-efficacy for exercise, and pain catastrophizing were found to be significant predictors of the change in disability from baseline to the 6-month follow-up (R^2 = 34.2%) (Table 15). These results

demonstrate that patients with a high pre-operative level of disability were more likely to have a larger reduction in postoperative disability than those with low preoperative disability. This relationship is illustrated in Figure 9. The model also indicated that patients with lower preoperative levels of self-efficacy for exercise and pain catastro-phizing tended to have a larger reduction of postoperative disability than those with higher preoperative levels.
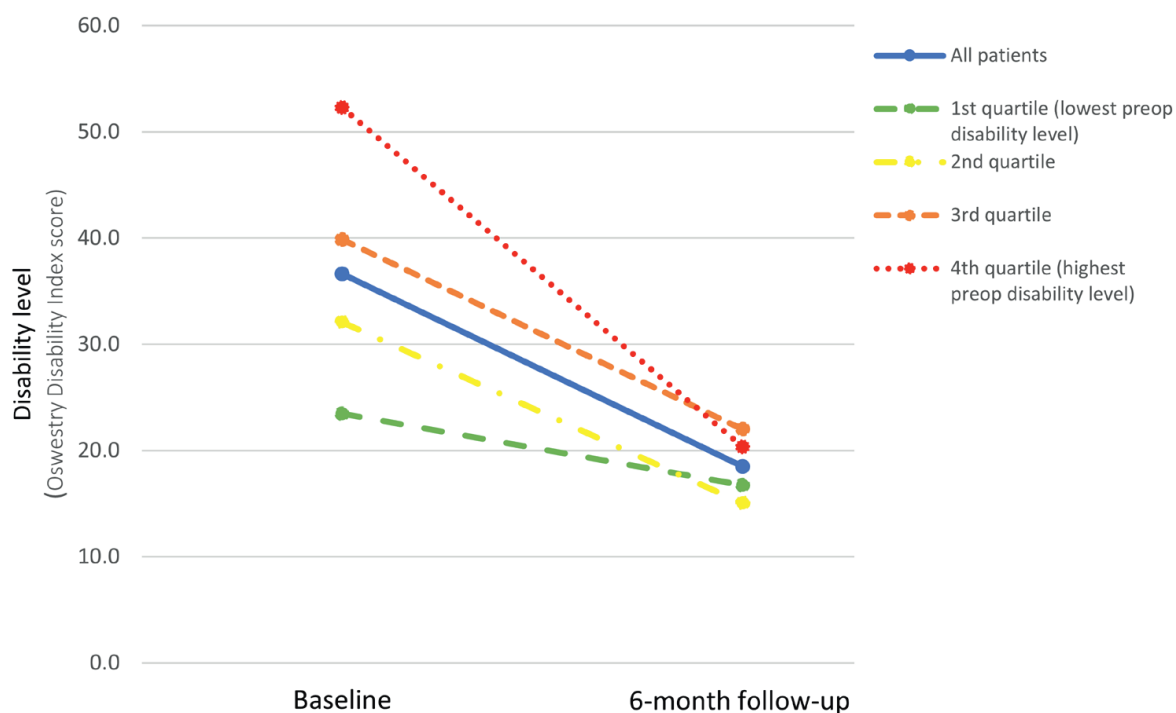

Figure 9. Scores at baseline and at 6-month follow-up for disability (Oswestry Disability Index, ODI), for all patients and for quartiles based on the patient's preoperative ODI level (Study IV). *Reproduced from World Neurosurgery, CC permission.*

DISCUSSION

The overall aim of the thesis was to investigate aspects of the measurement of functioning and physical activity in patients with LBP. Studies I and II concerned aspects of the measurement of functioning with physical capacity tasks and Studies III and IV concerned aspects of the measurement of physical activity with accelerometers.

5.1 MEASUREMENT OF FUNCTIONING WITH PHYSICAL CAPACITY TASKS

The first research gap identified in this thesis was that no previous study had made a synthesis of previous findings of reliability, validity, and responsiveness of physical capacity tasks for patients with LBP.

In Study I, the physical capacity tasks five-repetition sit-to-stand, five-minute walk, 50-foot walk, progressive isoinertial lifting evaluation, and timed up-and-go had the strongest evidence for adequate test-retest reliability and construct validity for the measurement of functioning in patients with chronic LBP. Of these, five-repetition sit-to-stand also showed adequate responsiveness. Of the other tasks in Study I that showed adequate responsiveness, one-minute stair climbing had the strongest evidence for adequate reliability. A number of tasks that involved walking were investigated specifically for patients with diagnoses known to severely affect walking capacity (see the level of evidence in Table 9 marked in *italics*). Of these tasks, timed up-and-go and shuttle walking were the only ones that showed adequate ratings for more than one measurement property.

During the course of Study I, a systematic review of the reliability of physical capacity tasks was published (Denteneer et al., 2018). The authors used parts of the COSMIN framework in their methods, but they did not, however, provide the level of evidence

for their findings, making comparisons with Study I more difficult. Nevertheless, a large part of the findings in Denteneer et al. are in line with reliability results of Study I—as five-repetition sit-to-stand, five-minute walking, and 50-foot walk had the best results for test-retest reliability.

In Study I, we concluded that five-repetition sit-to-stand, five-minute walk, 50-foot walk, progressive isoinertial lifting evaluation, timed up-and-go, and one-minute stair climbing were promising physical capacity tasks to measure functioning in patients with chronic LBP. The reason for labeling the tasks "promising," and not giving a final recommendation on which of the tasks could be used in research and clinical practice, was that more research on responsiveness and measurement error was needed. In Study II, we aimed to gain knowledge of the responsiveness of four of the promising tasks from Study I: five-minute walk, 50-foot walk, timed up-and-go, and 1-minute stair climbing. As Study II investigated the responsiveness of the tasks specifically for patients with chronic LBP due to DDD who underwent lumbar fusion surgery, I will give recommendations later in this discussion section on which tasks to use for measuring functioning for this patient group. However, I will first briefly summarize the results of Study II and relate them to previous research.

In Study II, we found that 50-foot walk, timed up-and-go, and one-minute stair climbing showed adequate responsiveness. In contrast, 5-minute walk showed inadequate responsiveness. In line with our results, Gautschi et al. (2016) found adequate responsiveness for timed up-and-go. That study comprised a mixed sample of patients with lumbar degenerative conditions undergoing lumbar spine surgery, with only a small proportion of patients with chronic LBP due to DDD. One other previous study has investigated the responsiveness of five-minute walk and one-minute stair climbing (Andersson et al., 2010) and two have investigated the responsiveness of 50-foot walk (Andersson et al., 2010; Strand et al., 2011). These studies concerned patients with chronic LBP who underwent non-surgical interventions. In line with Study II, Andersson et al. (2010) found that one-minute stair climbing had adequate responsiveness. Moreover, Andersson et al. found that five-minute walk had inadequate responsiveness. They reasoned that the finding might be a result of the possibility that the tasks are not challenging enough for patients with chronic LBP. Patients might therefore only show small improvements in this task after an intervention, which could limit the task's responsiveness. This reasoning is also in line with clinical experience regarding patients with chronic LBP due to DDD, which indicates that the patients do not usually have any major problems with walking on flat surfaces. Instead, walking on uneven surfaces is usually more challenging.

In contrast with Study II, Andersson et al. (2010) and Strand et al. (2011) found that 50-foot walk had inadequate responsiveness. This discrepancy may be due to dissimilarities in patient characteristics. Patients with chronic LBP due to DDD often have motion-elicited back pain, so that they can have difficulties with quick movements of the spine. As such, 50-foot walk could be challenging for these patients, as the task requires them to make a quick turn after having walked 25 feet. In contrast, the patients in the previous responsiveness studies (Andersson et al., 2010; Strand et al., 2011) may have found the task less challenging.

So, which physical capacity tasks are recommended for assessing functioning in patients with chronic LBP due to DDD who undergo lumbar fusion surgery? At this point, 50-foot walk and timed up-and-go can be recommended. First, the recommendations are mainly based on the results for responsiveness in Study II since the study is the only that has investigated the responsiveness of physical capacity tasks specifically for patients with chronic LBP due to DDD who have undergone lumbar fusion surgery. Second, the recommendations are supported by the findings for reliability and validity of 50-foot walk and timed up-and-go investigated for patients with chronic LBP in Study I:

- In addition to showing adequate responsiveness in Study II, 50-foot walk showed strong evidence for adequate test-retest reliability and moderate evidence for adequate construct validity in Study I.

- In addition to the adequate responsiveness seen in Study II, timed up-and-go demonstrated moderate evidence for adequate test-retest reliability and construct validity in Study I.

One-minute stair climbing was not included in the recommendations even though it demonstrated adequate responsiveness in Study II and moderate evidence for adequate reliability in Study I. The rationale for that decision was that the validity of the task has, to the best of my knowledge, not been investigated before. I would therefore suggest that future studies aim to investigate the validity of 1-minute stair climbing for patients with

chronic LBP due to DDD who undergo lumbar fusion surgery.

In Study II, we also investigated the MICs for one-minute stair climbing, 50-foot walk, timed up-and-go, and five-minute walk (Table 13). The MICs can be used to judge whether the postoperative change scores of the physical capacity tasks are perceived as important by patients (de Vet et al., 2006). In research, the MICs could, for example, be used to evaluate the proportion of "responders" to treatment, where patients with change scores larger than the MIC values are classified as responders (Guyatt et al., 1998). It is, however, important to acknowledge that the MIC is a group-based statistic and that the value for MIC might not always reflect an individual patient's view of the change (de Vet et al., 2010). Thus, when comparing an individual patient's change score in relation to the MIC in clinical practice, it is essential to interpret the change score in relation to the patient's reported experience of the change and not only the MIC. Comparing individual change scores with the MICs might, for instance, serve as a reference for what the "average" patient finds important and could possibly aid the shared decision-making process in the patient's postoperative rehabilitation. However, the 95% confidence intervals of the MICs were wide, and they should therefore be viewed with some caution.

In order to detect changes that are as small as the MIC, it is crucial that the latter is larger than the smallest detectable change (defined as "the smallest change that can be detected by the measurement instrument, beyond measurement error" (de Vet et al., 2011)). Of the recommended tasks, only one-minute stair climbing showed a measurement error that was smaller than the smallest detectable change observed in previous studies (Simmonds et al., 1998; Smeets et al., 2006). Thus, if patients show a change that is equal to the MIC of 20 steps, the healthcare professional can be fairly certain that the change is both important *and* statistically significant (de Vet et al., 2006). In contrast, for the 50-foot walk and timed up-and-go, the MICs were smaller than the smallest detectable change found in previous studies (Andersson et al., 2010; Gautschi et al., 2017). Consequently, when patients show a change that is equal to the MICs for these tasks, the change is most likely not statistically significant even though that change may be perceived as important by the patients (de Vet et al., 2006). The previously derived values for the smallest detectable change were, however, not investigated for patients with chronic LBP due to DDD scheduled for lumbar fusion surgery. As the smallest detectable change is context-dependent (C. B. Terwee et al., 2010), I encourage researchers to investigate the smallest detectable change specifically for this patient group.

5.2 MEASUREMENT OF PHYSICAL ACTIVITY WITH ACCELEROMETERS

The second research gap identified in this thesis was that no previous studies had used accelerometers to investigate the level of physical activity in patients with chronic LBP due to DDD who undergo lumbar fusion surgery.

Study III showed that patients with chronic LBP due to DDD had a low level of physical activity relative to the WHO recommendations (World Health Organization, 2009b), with only 17% of them fulfilling these recommendations. The patients' adherence to the WHO recommendations on physical activity is comparable to that found in

population-based studies in Norway (20%) (Hansen et al., 2012) and Germany (14–20%) (Krug et al., 2013; Luzak et al., 2017). Furthermore, the patients had a slightly lower level of physical activity than individuals in a Swedish population-based study (measured as minutes per day of at least moderate-intensity physical activity (non-bouted)) (Hagstromer et al., 2010) (Table 11).

Thus, at a group level, the patients' level of physical activity was similar to that found in population-based studies. However, a proportion of the patients had a very low level of physical activity. More specifically, 28% of the patients did not spend a single minute on at least moderate-intensity physical activity (in 10-minute bouts). If these patients were to continue with their low level of physical activity, they would have an increased risk of developing additional diseases such as diabetes, cardiovascular disease, and cancer (I. M. Lee et al., 2012; Wen et al., 2011; World Health Organization, 2009a).

So, how did the patients' level of physical activity change six months after surgery? As seen in the descriptive statistics in Study IV, the patients with the lowest level of preoperative physical activity (the 1st quartile in Figure 8) increased their level of physical activity by the greatest amount after surgery. However, the patients still had a low level of physical activity in compared with recommendations on physical activity and the general population in Sweden (Hagströmer et al., 2010). It therefore appears that these patients may need extra interventions to increase their level of physical activity further. The largest health effects of physical activity can be gained by encouraging those with the lowest level to be more physically active (World Health Organization, 2009a). By measuring the level of physical activity in clinical practice, the patients with the lowest level could, theoretically, be identified and then invited to participate in pre- or postoperative interventions aimed at increasing the patients' level of physical activity.

5.3 MEASUREMENT OF FEAR-AVOIDANCE VARIABLES TO IDENTIFY BARRIERS TO AND PREDICTORS OF HEALTH

The third research gap identified in this thesis was that no studies had investigated fear-avoidance variables as possible barriers to physical activity for patients with chronic LBP due to DDD who undergo lumbar fusion surgery.

In Study III, the preoperative levels of fear of movement and disability were negatively associated with the patient's preoperative level of physical activity (steps per day). This finding implies that patients with the highest levels of fear of movement and disability had the lowest preoperative levels of physical activity. Both fear of movement and disability have been shown to be modifiable variables in pre- and postoperative interven-

tions for patients with chronic LBP (Abbott et al., 2010a; Cabilan et al., 2016; Gilmore et al., 2015). Future pre- and postoperative interventions that target disability and fear of movement in patients with chronic LBP due to DDD might therefore increase the patients' levels of physical activity. However, no certain conclusions can be drawn on such a cause-effect relationship since Study III had a cross-sectional study design. Researchers are recommended to use longitudinal study designs to investigate further the role of fear of movement and disability in relation to the level of physical activity in patients with chronic LBP due to DDD who undergo lumbar fusion surgery.

The final research gap identified in the thesis was that there had been very little previous research on the value of fear-avoidance variables and walking capacity in predicting postoperative change in physical activity and disability in patients with chronic LBP due to DDD who undergo lumbar fusion surgery.

Study IV showed that a low preoperative level of physical activity and high preoperative level of self-efficacy for exercise were predictive of a more favorable postoperative outcome in physical activity. Furthermore, a high preoperative level of disability and low preoperative levels of pain catastrophizing and self-efficacy for exercise were found to be predictive of a more favorable outcome regarding disability. Walking capacity, as measured with five-minute walk and 50-foot walk, was not predictive of the postoperative outcomes in either physical activity or disability. The finding that a low preoperative level of pain catastrophizing could predict a more favorable postoperative outcome of disability is in line with a previous study on patients with chronic LBP undergoing lumbar fusion surgery (Abbott et al., 2011). By contrast, self-efficacy for exercise has, to my knowledge, not previously been investigated as a potential predictor of the postoperative outcome of lumbar spine surgery previously.

Thus, a number of fear-avoidance variables were found to be significant predictors of physical activity and disability. However, the fear avoidance variables accounted for only a limited proportion of the explained variance of the models. Nevertheless, even though a variable only adds a small proportion of the explained variance of a model, research suggests that it is still important to include it in the full prediction model, to yield better estimates (Steyerberg, 2009).

So, how could the prediction models be used in clinical practice? The explained variance of the physical activity model was low (25.1%), and research suggests that this value would probably be lower when applying the model in a context different from a clinical study (Steyerberg, 2009). Therefore, the model should, foremost, be viewed as a first step in the development of a more robust model. I therefore recommend researchers to investigate additional variables that could further increase the explained variance of the model. For instance, one promising variable is the intention to be more physically active, which has been shown to be one of the most robust predictors in population-based studies (Rhodes et al., 2015).

Nevertheless, despite the low explained variance, the prediction model for physical activity could possibly be used already at this point. A healthcare professional might use it to get a rough estimate of the postoperative level of physical activity in patients with chronic LBP due to DDD who are scheduled for lumbar fusion surgery. This information might help the healthcare professional to personalize pre- and postoperative interventions for a patient. As an example, a healthcare professional might want to provide more intensive pre- and postoperative interventions for patients who are predicted to have a low level of physical activity after surgery. However, as the prediction model provides only a rough estimate, the healthcare professional should obviously use his/her clinical experience regarding the postoperative outcome—and not only the predicted estimate.

An important distinction is that the prediction model for physical activity concerns the *change* in postoperative outcomes, and not the actual level of physical activity or disability six months postoperatively. For instance, a patient who demonstrates a significant postoperative increase in physical activity may still have a low level of physical activity six months after surgery. Consequently,

when using the prediction model of physical activity, the predicted estimate should be interpreted in relation to the patient's preoperative level of physical activity, as shown in Equation 1. The intercept of the prediction model was not included in the equation as it was rounded off to zero.

$$\text{Postop PA} = [\text{Preop PA}] - 0.349 \times [\text{preop PA}] + 0.176 \times [\text{preop SEES}]$$

Equation 1. Equation for estimating the 6-month postoperative physical activity measured as minutes per day of at least moderate-intensity physical activity (non-bouted)

PA, physical activity measured as minutes per day of at least moderate-intensity physical activity (non-bouted) with the GT3X+ accelerometer; SEES, total score on the Self-Efficacy for Exercise Scale

What, then, are the implications of the findings from the prediction model of disability? The explained variance of the prediction model of disability (34.2%) was comparable to that of other prediction models for patients undergoing lumbar fusion surgery (25.0–41.6%) (Abbott et al., 2011; Ekman et al., 2009; Trief et al., 2006). However, there are concerns regarding the contradictory results for self-efficacy for exercise. The result that a higher preoperative self-efficacy for exercise was predictive of a worse result for disability is namely opposite to what would be expected in relation to the modified fear-avoidance model (Lotzke et al., 2016). An explanation might be that the fear-avoidance model by Woby et al. (2007), on which Lotzke et al. (2016) based their model, included *functional* self-efficacy whereas we used a scale that measured self-efficacy for exercise. It is possible that self-efficacy for exercise might be a construct that is too unrelated to functional self-efficacy, and therefore also too unrelated to disability. Researchers should certainly be careful when using a prediction model if a variable predicts an outcome in the opposite direction to what is expected (Steyerberg, 2009). At this point, the findings regarding the predictors of the disability model are therefore recommended to be used to guide future research rather than be used in clinical practice. First, researchers investigating the prediction of postoperative disability are encouraged to include pain catastrophizing as a potential predictor. Second, functional self-efficacy is suggested to be a better candidate for prediction of postoperative disability than self-efficacy for exercise.

5.4 METHODOLOGICAL CONSIDERATIONS

5.4.1 Internal validity

Internal validity refers to the degree to which the results of a study are attributable to the independent variables included, and not to some rival explanation (Shadish et al., 2002).

Internal validity is closely connected to the quality of outcome measures (de Vet et al., 2011; Shadish et al., 2002). The use of high-quality outcome measures with support for reliability and validity strengthens the internal validity of the thesis. In addition, most of the outcome measures were investigated for reliability and validity specifically for patients with LBP.

Study I followed the recommended procedure of the COSMIN initiative for conducting a systematic review. In addition to this procedure, a third author was consulted to resolve discrepancies in the screening of abstracts, selection of full-text articles, and

assessment of methodological quality. We believe that this additional methodological precaution strengthened the internal validity of the study. Moreover, the author who resolved discrepancies in the assessment of methodological quality (L.B. Mokkink) is one of the developers of the COSMIN 4-point checklist, and she made sure that we used the checklist appropriately. There are also threats to the internal validity of Study I. First, possible publication bias may have affected the results, as unpublished studies are more likely to report unfavorable results (Dwan et al., 2013). As there are, no registries for studies of reliability, validity, and responsiveness as there are for RCTs, we did not perform any analysis of publication bias in our review. Second, the COSMIN 4-point checklist for assessment of the methodological quality was developed for PROMs (Mokkink et al., 2010b; C. Terwee et al., 2012) and not for physical capacity tasks. However, the authors of the checklist stated that they phrased the items of the checklist in a general manner, and that it could be used for types of measurement instruments other than PROMs, including physical capacity tasks (Mokkink et al., 2010b). Furthermore, to our knowledge, the checklist is the only consensus-based quality assessment tool for studies on measurement properties, so it appears to have been the best available quality assessment tool for Study I.

The use of accelerometers in the measurement of physical activity strengthened the internal validity of Studies III and IV. Accelerometers have been shown to yield less biased estimates of the level of physical activity than physical activity PROMs. More specifically, researchers have recommended accelerometers over physical activity PROMs, since they are not reliant on accurate recall of the intensity, frequency, and duration of daily activities—and they appear to be less subject to overestimations and social desirability (Cerin

et al., 2016; Prince et al., 2008; Slootmaker et al., 2009). However, measurement of physical activity with accelerometers has certain limitations. First, accelerometers worn on the hip mainly measure ambulatory activities, and cannot accurately record activities that cause little movement of the body's center of gravity, such as cycling and weightlifting. However, research suggests that walking is the major contributor to the individual's level of physical activity (Hagströmer et al., 2006), and I therefore believe that the accelerometer provided adequate estimations regarding most of the patients in our studies. Second, assessment of physical activity with an accelerometer relies on the assumption that the wear period (seven days in Studies II and IV) reflects the patient's habitual physical activity. Nevertheless, wearing an accelerometer may have motivated the patients to be more physically active, a phenomenon termed "reactivity" (Baumann et al., 2018; Davis et al., 2016). Despite these limitations, research suggests that accelerometers provide relatively unbiased estimates of the level of physical activity compared with physical activity PROMs (Cerin et al., 2016; Prince et al., 2008; Slootmaker et al., 2009).

In Study III, the level of physical activity was measured as minutes per day of at least moderate-intensity physical activity performed in bouts of at least 10 minutes. Researchers have, however, raised concerns about the 10-minute bout criterion. One concern is that certain types of physical activity, such as interval training, are excluded from the measurement of physical activity using this criterion (2018 Physical Activity Guidelines Advisory Committee, 2018). The 10-minute bout criterion may therefore lead to underestimation of an individual's level of physical activity, including that of the patients in this thesis. In line with this reasoning, recently updated recommendations on physical activity have removed the 10-minute bout crite-

rion (2018 Physical Activity Advisory Committee, 2018). However, as Study III compared the levels of physical activity in patients with the *WHO recommendations* on physical activity, we argue that using the 10-minute bout criterion was appropriate.

The explained variance of the regression models of physical activity in Studies III and IV was low. This indicates that there were other important factors concerning physical activity that the models did not cover (Fahrmeir et al., 2013). So-called multilevel ecological models of physical activity have shown good results in explaining the variance in physical activity in population-based studies. These models include intrapersonal, interpersonal, environmental, regional, national, and global factors (Bauman et al., 2012; Sallis et al., 2008). One possible threat to the internal validity of Studies III and IV is that we included only intrapersonal variables and not other factors from the multilevel ecological models. However, the purpose of Studies III and IV was above all to investigate associations between the level of physical activity and fear-avoidance variables. Adding additional factors from multilevel ecological models, such as the accessibility to recreation facilities and social support (Baumann et al., 2018; Choi et al., 2017; Sallis et al., 2008), would nevertheless have strengthened the internal validity of the studies, and these should preferably be included in future studies using larger sample sizes.

Studies III and IV used a purposeful selection method for regression analysis that was developed by Hosmer et al. (1999) and further developed by Bursac et al. (2008). The method is one of many so-called "stepwise" methods for regression analysis, which aim to increase the statistical power of the analysis by removing independent variables that are insufficiently associated with the dependent variable (Fahrmeir et al., 2013). The method

in Studies III and IV was chosen due to having a few benefits over traditional stepwise methods. First, the threshold for the p-value in the initial univariate analysis (0.25) was higher than in traditional stepwise methods (such as 0.05 or 0.10) (Fahrmeir et al., 2013). When the threshold for the p-value is high, the risk of wrongly excluding important independent variables will be lower than when using lower thresholds (Bursac et al., 2008; Fahrmeir et al., 2013). Second, in the last step of the method, independent variables that were excluded in the univariate analysis were added back to the multiple regression model, one by one, to see if they were significantly associated with the dependent variable. This step was performed to identify independent variables that were associated with the dependent variable only in the presence of other independent variables but not to the dependent variable alone. Traditional stepwise methods do not usually include this step (Fahrmeir et al., 2013) and they may thereby fail to identify important independent variables (Steyerberg, 2009).

5.4.2 External validity

The external validity denotes the extent to which the results of a study can be generalized (Shadish et al., 2002).

A representative sample is required if findings are to be generalized beyond the research setting. Patients with LBP are a heterogeneous patient group, and the eligibility criteria of the studies in the thesis were chosen to reduce heterogeneity and potential confounding bias. Even so, the patients in Study I were a heterogeneous sample, which is partly reflected in the differences in levels of pain and disability seen in the articles included. Although I acknowledge that such heterogeneity can influence the results of physical capacity testing (e.g. the number of meters walked or the amount of weight lifted), there appears to be little evidence that

the heterogeneity would affect the results of *measurement properties.* This is not least seen in the almost unanimous positive results for test-retest reliability and construct validity, regardless of the characteristics of the study populations. A smaller proportion of the results from Study I is, however, generalizable mainly to patients with back-related diagnoses that are known to severely affect walking capacity, such as lumbar spinal stenosis (see the level of evidence in *italics* in Table 9).

The patients in Studies II–IV had similar characteristics to those of patients with LBP due to DDD registered in Swespine in terms of age, duration of symptoms, and the relative proportion of men and women (Fritzell et al., 2018). This observation strengthens the external validity of Studies II–IV. However, there is reason to believe that the results in Studies II–IV are most likely *not* generalizable to all patients with LBP due to DDD who undergo lumbar fusion surgery. First, selection bias may limit the generalizability. The patients were part of an RCT that required them to travel to one of the spine clinics to see a physiotherapist on four occasions before surgery (Lotzke et al., 2016). Patients with higher preoperative levels of disability and pain intensity may therefore have declined participation in the study due to the probable burden of traveling to the clinic. This assumption is supported by the fact that the study sample reported having lower preoperative levels of disability and intensity of back pain compared to patients in Swespine who were scheduled for lumbar fusion surgery for chronic LBP due to DDD (Fritzell et al., 2018). Moreover, the proportion of patients undergoing instrumented posterior fusion of 1–2 segments instead of more invasive types of lumbar fusion procedures was higher than in Swespine.

Second, attrition bias may also have limited the external validity of the results of Studies II and IV. Patients who were lost to follow-up in the responsiveness analysis in Study II (n = 25) reported having significantly higher preoperative levels of depression, fear of movement, and pain catastrophizing than did the patients who were included in the responsiveness analysis. Moreover, lost to follow-up in the physical activity analysis in Study IV (n = 28) had a significantly higher preoperative level of back pain intensity and a significantly lower preoperative level of physical activity than did the patients who were included in the prediction analysis. Patients who are lost to follow-up reporting worse preoperative scores is nevertheless a common phenomenon seen in orthopedic clinical studies (Somerson et al., 2016).

Third, half of the sample in Studies II–IV was randomized to participate in a prehabilitation program before surgery (Lotzke et al., 2016). The participation in the prehabilitation program might have affected the patients' postoperative outcomes and thereby reduced the external validity of the results in Studies II–IV. There were, however, no statistically significant between-group differences seen in the RCT at follow-up, except for health-related quality of life (Lotzke et al., 2018). Based on these results, the participation in the prehabilitation program appears to have been a minor threat to the external validity.

5.5 ETHICAL CONSIDERATIONS

Even though ethical approval is not needed for systematic reviews, there are still ethical considerations. First, studies included in a systematic review could have some ethical

insufficiencies, and re-publishing the results from such studies in a systematic review would be questionable from an ethical point of view (Vergnes et al., 2010). However, we did not encounter any methodology in the studies that appeared to violate the Declaration of Helsinki. Second, studies included in a systematic review may have conflicts of interests, and authors of systematic reviews should therefore try to report them (Vergnes et al., 2010). To our knowledge, there were no conflicts of interests in the included studies for us to report.

According to the principle of individual autonomy, participants in a study need to be provided with sufficient information to make an informed decision about participation in a research project. All the patients in Studies II–IV received oral and written information about the purpose and the procedure of the project, potential risks and benefits regarding participation, data handling, and procedures for preservation of privacy. Moreover, the patients were informed that they could end participation at any moment without giving any reason for doing so. All the patients who participated in the projects signed an informed consent form after having received the written and oral information about the project.

A potential risk regarding the physical capacity tasks was that performing the tasks might have caused the patients more back pain. However, the physical capacity tasks included only activities commonly performed in daily life (walking, climbing stairs, and rising up from a chair), and we therefore considered the risk of sustained pain to be small. Moreover, the physical capacity tasks were performed at a spine clinic with experienced healthcare professionals who could be of help in the unlikely event of sustained pain. There was, however, no need for this—as there were no adverse events related to performing the physical capacity tasks. Another potential risk of participating in the project was the time it took the patients to visit one of the spine clinics for the follow-up assessments (3, 6, 12, and 24 months after surgery). However, visiting the clinic as frequently as this might also be considered a benefit. The routine postoperative assessment of patients in clinical practice only occurs six months after surgery, and the follow-up assessments for the project would possibly identify problems related to the surgery that would otherwise have gone unnoticed.

The procedures for the collection and handling of data were designed from an ethical point of view. Regarding the collection of data, the set of PROMs was kept as short as possible while still including all relevant data, and the booking of appointments for follow-up was performed according to a protocol that stipulated the maximum number of attempts for reaching a patient. The procedure for the handling of data was designed to keep data safe and confidential. Electronic data were stored on a password-protected computer and routinely backed up. Data from PROMs were stored in a locked, fireproof cupboard. All data were coded, and the code key was stored separately from the data.

FUTURE RESEARCH

Physical capacity tasks and accelerometers are not routinely used to evaluate the outcome of lumbar spine surgery, although these outcome measures can give knowledge of a patient's health that disability PROMs do not (Gautschi et al., 2016b; Lin et al., 2011; Simmonds et al., 1998). It is desirable that future studies should investigate the feasibility and cost-effectiveness of implementing the use of physical capacity tasks and accelerometers in clinical practice.

I can also see the benefits of developing a "health battery" made up of several types of outcome measures, including physical capacity tasks, disability PROMs, and measurements of physical activity. Using such a battery would possibly facilitate shared decision-making with the patient, to focus the intervention on the aspects of health that are most important. This type of battery could be partly based on the results of reliability, validity, and responsiveness in Studies I and II. It would be important to assess the content validity of such a battery, since this measurement property was not assessed in any of the articles included in Study I and also because the measurement property has hardly been investigated for disability PROMs (Chiarotto et al., 2017).

Furthermore, I encourage researchers to investigate the postoperative level of physical activity in patients with LBP due to DDD for follow-up periods greater than six months. Results from studies with, for example, 1-year and 2-year follow-up might provide knowledge of when patients tend to reach a plateau in changes of physical activity. Such knowledge could be used to determine the optimal time point for measuring the postoperative level of physical activity in patients who have undergone lumbar fusion surgery.

I would also recommend that future studies should aim to measure the level of physical activity in a spinal orthopedic surgical context from a broader perspective. First, more knowledge is needed on the pre- and postoperative level of physical activity in patients with lumbar spinal stenosis. Patients with lumbar spinal stenosis comprise the largest patient group undergoing elective lumbar spine surgery in Sweden (Fritzell et al., 2018), but little is known of their level of physical activity. Second, more knowledge is required of the physical activity of patients with chronic LBP who are denied surgery due to poor health. Clinical experience indicates that many of these individuals have a very low level of physical activity and are in need of interventions in order to be more physically active, but there has been little or no research on this matter.

It is desirable that future research should continue to develop prediction models of physical activity and disability. First, researchers should be encouraged to investigate the prediction model of physical activity in this thesis in a different setting, thereby testing its external validity (Seel et al., 2012; Steyerberg, 2009). Second, based on the results of Study IV, preoperative self-efficacy for exercise appears to be a promising predictor to consider in future prediction models of physical activity. Third, based on the results of Study IV, preoperative pain catastrophizing would appear to be a promising predictor when using future prediction models of disability.

Finally, I would also recommend that future studies investigate to use longitudinal study designs to investigate further the role of fear of movement and disability in relation to the level of physical activity in patients with chronic LBP due to DDD who undergo lumbar fusion surgery. These variables could be important to target in pre- and postoperative interventions aimed at increasing the patient's level of physical activity.

CONCLUSIONS

- Fifty-foot walk and timed up-and-go showed adequate results for reliability, validity, and responsiveness and are recommended for measurement of functioning in patients with chronic LBP with DDD who undergo lumbar fusion surgery.

- Patients with LBP due to DDD who had the highest levels of fear of movement and disability had the lowest preoperative level of physical activity. Future pre- and postoperative interventions targeting fear of movement and disability might increase the level of physical activity in patients who have a low preoperative level of physical activity.

- Patients with LBP due to DDD with a low preoperative level of physical activity were more likely to increase their level of physical activity after lumbar fusion surgery than those with higher preoperative levels. However, the patients with a low preoperative level of physical activity still had a low level of physical activity after surgery, and may therefore need extra pre- and postoperative interventions to reach a higher level.

- The prediction model of physical activity in this thesis could possibly aid healthcare professionals to—already before surgery—identify which patients are in need of extra pre- and postoperative interventions in order to reach a higher level of physical activity after lumbar fusion surgery.

ACKNOWLEDGEMENTS

There are many individuals who have contributed to this thesis in one way or another. I would especially like to thank:

All patients who were included in the project: thank you for your participation, not least for filling out the same questionnaires over and over again—this thesis would not have been possible without you!

Mari Lundberg, main supervisor: my sincere thanks for your boundless devotion in guiding, supporting, and challenging me during my PhD years. Thank you for believing in me from the start and giving me the opportunity to pursue my academic goals. Thank you also for exceeding your role as a supervisor through your support when times were rough in my private life. You are the most dedicated person I know—never let go of your quest for knowledge (and food)!

Annelie Gutke, assistant supervisor: thank you for your unwavering calmness and your skill of always seeing things from new perspectives and conveying these perspectives in such a pedagogic way. You are a role model when it comes to teaching!

Rob Smeets, assistant supervisor: thank you for your analytic skills, precise feedback, and for stimulating my thinking. Thank you also for introducing me to Tacx bike trainers—buying one and using it helped to preserve at least some of my physical and mental health during the last year of my studies!

Helena Brisby, assistant supervisor: thank you for your unending positivity and for your feedback and warm support when I have needed it the most. Your thoughtfulness towards your students is heart-warming!

Hanna Lotzke, fellow PhD student: thank you for walking this journey alongside me. Thank you for sharing your project and knowledge so generously and for always taking the time to listen to me in times of distress. I will always fix your computer with joy!

Maria Hagströmer, co-author: thank you for your enthusiasm for teaching the art of measuring physical activity—you are an inexhaustible source of inspiration!

Lidwine Mokkink, co-author: thank you for your expertise and patience when answering the hundreds (thousands?) of questions on measurement properties over the years.

Olle Hägg, co-author: thank you for all the valuable feedback over the years—you have a fabulous skill to interpret and perform research from a clinical point of view, which I admire.

All colleagues in Back in Motion Research Group: thank you for all the support, feedback, and fika during the years. It is truly an honor to be a part of your group!

Christer Söderling, physiotherapist colleague: thank you for your patient and meticulous work in the project and for always raising the mood with your humor, which is somehow both the best and the worst.

All colleagues at Spine Center Gothenburg: thank you for your valuable help in the project, the wonderful atmosphere, and for all the cheese!

All colleagues at the Lundberg Laboratory for Orthopaedic Research at Mölndal Hospital: thank you for all the encouragement, feedback, and laughs over the years.

All colleagues at the Division of Home Rehabilitation, Department of Nursing and Care for the Elderly in Borås Stad: thank you for all your support and for being so wonderful colloeagues!

All colleagues at the Department of Rehabilitation Medicine at Maastricht University: thank you for accommodating me so warmly during my month in Maastricht.

Anna Ekman and Mattias Sundén, statisticians: thank you for everything that you have taught me and for inspiring me to learn more statistics.

Alistair Kidd and Les Hearn, language editors: thank you for your attention to detail and prompt work in the editing of the thesis.

Cina Holmer, administrator: thank you for your kindness when taking care of all the paper work surrounding the thesis.

Kjell, Anita, Malin, Viktor, Daphne, and Jonathan: thank you for taking care of Oliver as if he were your own during intense periods of work. It takes time, space, and sanity to complete any profound scientific work, all of which are luxuries in life with a toddler. I am so grateful for that you provided me with those luxuries during my studies.

Finally, my deepest gratitude, that I could never express in words, is for my family:

My parents Eva and Bengt, my brother Mattias and his family, and my grandmother Barbro: thank you for your boundless love and for always believing and supporting my educational goals and all other endeavors in life. You are the most amazing family members anyone could wish for!

Above all, *my son Oliver and my wife Marika:* Oliver, thank you for constantly reminding me of what is most important in life (loving, laughing, and playing): Marika, thank you for bringing so much joy into my life and for being so patient with my late evenings, early mornings, busy weekends, and canceled vacations during the last years. Literally, I would not have made it without you. I Love You!

REFERENCES

2018 Physical Activity Guidelines Advisory Committee. (2018). The 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Retrieved from https://health.gov/paguidelines/second-edition/report/pdf/PAG_Advisory_Committee_Report.pdf

Abbott, A. D., Tyni-Lenne, R., & Hedlund, R. (2011). Leg pain and psychological variables predict outcome 2-3 years after lumbar fusion surgery. *Eur. Spine J., 20*(10), 1626-1634. doi:10.1007/s00586-011-1709-6

Abbott, A. D., Tyni-Lenné, R., & Hedlund, R. (2010a). Early rehabilitation targeting cognition, behavior, and motor function after lumbar fusion: A randomized controlled trial. *Spine (Phila Pa 1976), 35*(8), 848-857.

Abbott, A. D., Tyni-Lenné, R., & Hedlund, R. (2010b). The influence of psychological factors on pre-operative levels of pain intensity, disability and health-related quality of life in lumbar spinal fusion surgery patients. *Physiotherapy, 96*(3), 213-221. doi:http://dx.doi.org/10.1016/j.physio.2009.11.013

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Jr., Tudor-Locke, C., Greer, J. L., Vezina, J., Whitt-Glover, M. C., & Leon, A. S. (2011). 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med. Sci. Sports Exerc., 43*(8), 1575-1581. doi:10.1249/MSS.0b013e31821ece12

Airaksinen, O., Brox, J. I., Cedraschi, C., Hildebrandt, J., Klaber-Moffett, J., Kovacs, F., Mannion, A. F., Reis, S., Staal, J. B., Ursin, H., & Zanoli, G. (2006). Chapter 4. European guidelines for the management of chronic nonspecific low back pain. *Eur. Spine J., 15 Suppl 2*, S192-300. doi:10.1007/s00586-006-1072-1

Andersen, T., Christensen, F. B., Laursen, M., Hoy, K., Hansen, E. S., & Bunger, C. (2001). Smoking as a predictor of negative outcome in lumbar spinal fusion. *Spine (Phila Pa 1976), 26*(23), 2623-2628.

Andersson, E. I., Lin, C. C., & Smeets, R. J. (2010). Performance tests in people with chronic low back pain: responsiveness and minimal clinically important change. *Spine (Phila Pa 1976), 35*(26), E1559-1563. doi:10.1097/BRS.0b013e3181cea12e

Armstrong, M., McDonough, S., & Baxter, D. (2005). Reliability and repeatability of the shuttle walk test in patients with chronic low back pain. *International Journal of Therapy and Rehabilitation, 12*(10), 438-443. doi:10.12968/ijtr.2005.12.10.24531

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev., 84*(2), 191-215.

Bauman, A. E., Reis, R. S., Sallis, J. F., Wells, J. C., Loos, R. J., & Martin, B. W. (2012). Correlates of physical activity: why are some people physically active and others not? *Lancet, 380*(9838), 258-271. doi:10.1016/s0140-6736(12)60735-1

Baumann, S., Gross, S., Voigt, L., Ullrich, A., Weymar, F., Schwaneberg, T., Dorr, M., Meyer, C., John, U., & Ulbricht, S. (2018). Pitfalls in accelerometer-based measurement of physical activity: The presence of reactivity in an adult population. *Scand. J. Med. Sci. Sports, 28*(3), 1056-1063. doi:10.1111/sms.12977

Bernstein, I. A., Malik, Q., Carville, S., & Ward, S. (2017). Low back pain and sciatica: summary of NICE guidance. *BMJ, 356*. doi:10.1136/bmj.i6748

Blain, H., Carriere, I., Sourial, N., Berard, C., Favier, F., Colvez, A., & Bergman, H. (2010). Balance and walking speed predict subsequent 8-year mortality independently of current and intermediate events in well-functioning women aged 75 years and older. *J. Nutr. Health Aging, 14*(7), 595-600.

Brakenhoff, T. B., Mitroiu, M., Keogh, R. H., Moons, K. G. M., Groenwold, R. H. H., & van

Smeden, M. (2018a). Measurement error is often neglected in medical literature: a systematic review. *J. Clin. Epidemiol.* doi:10.1016/j.jclinepi.2018.02.023

Brakenhoff, T. B., van Smeden, M., Visseren, F. L. J., & Groenwold, R. H. H. (2018b). Random measurement error: Why worry? An example of cardiovascular risk factors. *PLoS One, 13*(2), e0192298. doi:10.1371/journal.pone.0192298

Brodke, D. S., Goz, V., Lawrence, B. D., Spiker, W. R., Neese, A., & Hung, M. (2017). Oswestry Disability Index: a psychometric analysis with 1,610 patients. *Spine J, 17*(3), 321-327. doi:10.1016/j.spinee.2016.09.020

Brox, J. I., Sørensen, R., Friis, A., Nygaard, Ø., Indahl, A., Keller, A., Ingebrigtsen, T., Eriksen, H. R., Holm, I., Koller, A. K., Riise, R., & Reikerås, O. (2003). Randomized Clinical Trial of Lumbar Instrumented Fusion and Cognitive Intervention and Exercises in Patients with Chronic Low Back Pain and Disc Degeneration. *Spine (Phila Pa 1976), 28*(17), 1913-1921. doi:10.1097/01.brs.0000083234.62751.7a

Buchbinder, R., van Tulder, M., Öberg, B., Costa, L. M., Woolf, A., Schoene, M., Croft, P., Buchbinder, R., Hartvigsen, J., Cherkin, D., Foster, N. E., Maher, C. G., Underwood, M., van Tulder, M., Anema, J. R., Chou, R., Cohen, S. P., Menezes Costa, L., Croft, P., Ferreira, M., Ferreira, P. H., Fritz, J. M., Genevay, S., Gross, D. P., Hancock, M. J., Hoy, D., Karppinen, J., Koes, B. W., Kongsted, A., Louw, Q., Öberg, B., Peul, W. C., Pransky, G., Schoene, M., Sieper, J., Smeets, R. J., Turner, J. A., & Woolf, A. (2018). Low back pain: a call for action. *The Lancet, 391*(10137), 2384-2388. doi:10.1016/S0140-6736(18)30488-4

Bursac, Z., Gauss, C. H., Williams, D. K., & Hosmer, D. W. (2008). Purposeful selection of variables in logistic regression. *Source Code Biol. Med., 3*, 17. doi:10.1186/1751-0473-3-17

Bydon, M., De la Garza-Ramos, R., Macki, M., Baker, A., Gokaslan, A. K., & Bydon, A. (2014). Lumbar fusion versus nonoperative management for treatment of discogenic low back pain: a sys-

tematic review and meta-analysis of randomized controlled trials. *J Spinal Disord Tech, 27*(5), 297-304. doi:10.1097/bsd.0000000000000072

Cabilan, C. J., Hines, S., & Munday, J. (2016). The Impact of Prehabilitation on Postoperative Functional Status, Healthcare Utilization, Pain, and Quality of Life: A Systematic Review. *Orthop. Nurs., 35*(4), 224-237. doi:10.1097/nor.0000000000000264

Campbell, H., Rivero-Arias, O., Johnston, K., Gray, A., Fairbank, J., & Frost, H. (2006). Responsiveness of objective, disease-specific, and generic outcome measures in patients with chronic low back pain: an assessment for improving, stable, and deteriorating patients. *Spine (Phila Pa 1976), 31*(7), 815-822. doi:10.1097/01.brs.0000207257.64215.03

Carlsson, A. M. (1983). Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. *Pain, 16*(1), 87-101.

Carreon, L. Y., Glassman, S. D., Djurasovic, M., Dimar, J. R., Johnson, J. R., Puno, R. M., & Campbell, M. J. (2009). Are preoperative health-related quality of life scores predictive of clinical outcomes after lumbar fusion? *Spine (Phila Pa 1976), 34*(7), 725-730. doi:10.1097/BRS.0b013e318198cae4

Carvalho, F. A., Maher, C. G., Franco, M. R., Morelhao, P. K., Oliveira, C. B., Silva, F. G., & Pinto, R. Z. (2017). Fear of Movement Is Not Associated With Objective and Subjective Physical Activity Levels in Chronic Nonspecific Low Back Pain. *Arch. Phys. Med. Rehabil., 98*(1), 96-104. doi:10.1016/j.apmr.2016.09.115

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep., 100*(2), 126-131.

Cerin, E., Cain, K. L., Oyeyemi, A. L., Owen, N., Conway, T. L., Cochrane, T., D, V. A. N. D., Schipperijn, J., Mitas, J., Toftager, M., Aguinaga-Ontoso, I., & Sallis, J. F. (2016). Correlates of Agreement between Accelerometry and Self-reported Phys-

ical Activity. *Med. Sci. Sports Exerc., 48*(6), 1075-1084. doi:10.1249/mss.0000000000000870

Chapman, J. R., Norvell, D. C., Hermsmeyer, J. T., Bransford, R. J., DeVine, J., McGirt, M. J., & Lee, M. J. (2011). Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine (Phila Pa 1976), 36*(21 Suppl), S54-68. doi:10.1097/BRS.0b013e31822ef74d

Chen, K. Y., & Bassett, D. R., Jr. (2005). The technology of accelerometry-based activity monitors: current and future. *Med. Sci. Sports Exerc., 37*(11 Suppl), S490-500.

Chiarotto, A., Boers, M., Deyo, R. A., Buchbinder, R., Corbin, T. P., Costa, L. O. P., Foster, N. E., Grotle, M., Koes, B. W., Kovacs, F. M., Christine Lin, C. W., Maher, C. G., Pearson, A. M., Peul, W. C., Schoene, M. L., Turk, D. C., van Tulder, M. W., Terwee, C. B., & Ostelo, R. W. (2018). Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain, 159*(3), 481-495. doi:10.1097/j.pain.0000000000001117

Chiarotto, A., Maxwell, L. J., Terwee, C. B., Wells, G. A., Tugwell, P., & Ostelo, R. W. (2016). Roland-Morris Disability Questionnaire and Oswestry Disability Index: Which Has Better Measurement Properties for Measuring Physical Functioning in Nonspecific Low Back Pain? Systematic Review and Meta-Analysis. *Phys. Ther., 96*(10), 1620-1637. doi:10.2522/ptj.20150420

Chiarotto, A., Ostelo, R. W., Boers, M., & Terwee, C. B. (2017). A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in low back pain. *J. Clin. Epidemiol.* doi:10.1016/j.jclinepi.2017.11.005

Choi, J., Lee, M., Lee, J. K., Kang, D., & Choi, J. Y. (2017). Correlates associated with participation in physical activity among adults: a systematic review of reviews and update. *BMC Public Health, 17*(1), 356. doi:10.1186/s12889-017-4255-2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: L. Erlbaum Associates.

Conway, J., Tomkins, C. C., & Haig, A. J. (2011). Walking assessment in people with lumbar spinal stenosis: capacity, performance, and self-report measures. *Spine J, 11*(9), 816-823. doi:10.1016/j.spinee.2010.10.019

Davis, R. E., & Loprinzi, P. D. (2016). Examination of Accelerometer Reactivity Among a Population Sample of Children, Adolescents, and Adults. *J Phys Act Health, 13*(12), 1325-1332. doi:10.1123/jpah.2015-0703

de Schepper, E. I., Damen, J., van Meurs, J. B., Ginai, A. Z., Popham, M., Hofman, A., Koes, B. W., & Bierma-Zeinstra, S. M. (2010). The association between lumbar disc degeneration and low back pain: the influence of age, gender, and individual radiographic features. *Spine (Phila Pa 1976), 35*(5), 531-536. doi:10.1097/BRS.0b013e3181aa5b33

de Vet, H. C., Terluin, B., Knol, D. L., Roorda, L. D., Mokkink, L. B., Ostelo, R. W., Hendriks, E. J., Bouter, L. M., & Terwee, C. B. (2010). Three ways to quantify uncertainty in individually applied "minimally important change" values. *J. Clin. Epidemiol., 63*(1), 37-45. doi:10.1016/j.jclinepi.2009.03.011

de Vet, H. C., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: a practical guide*. Cambridge: Cambridge University Press.

de Vet, H. C., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M. (2006). Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes, 4*, 54-54. doi:10.1186/1477-7525-4-54

de Winter, J. C., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychol. Methods, 21*(3), 273-290. doi:10.1037/met0000079

DeBerard, M. S., Masters, K. S., Colledge, A. L., & Holmes, E. B. (2003). Presurgical biopsychosocial

---

variables predict medical and compensation costs of lumbar fusion in Utah workers' compensation patients. *Spine J, 3*(6), 420-429.

Deen, H. G., Jr., Zimmerman, R. S., Lyons, M. K., McPhee, M. C., Verheijde, J. L., & Lemens, S. M. (2000). Test-retest reproducibility of the exercise treadmill examination in lumbar spinal stenosis. *Mayo Clin. Proc., 75*(10), 1002-1007.

den Boer, J. J., Oostendorp, R. A., Beems, T., Munneke, M., & Evers, A. W. (2006). Continued disability and pain after lumbar disc surgery: the role of cognitive-behavioral factors. *Pain, 123*(1-2), 45-52. doi:10.1016/j.pain.2006.02.008

Denison, E., Åsenlöf, P., & Lindberg, P. (2004). Self-efficacy, fear avoidance, and pain intensity as predictors of disability in subacute and chronic musculoskeletal pain patients in primary health care. *Pain, 111*(3), 245-252. doi:10.1016/j.pain.2004.07.001

Denteneer, L., Van Daele, U., Truijen, S., De Hertogh, W., Meirte, J., & Stassijns, G. (2018). Reliability of physical functioning tests in patients with low back pain: a systematic review. *Spine J, 18*(1), 190-207. doi:10.1016/j.spinee.2017.08.257

Deyo, R. A. (1988). Measuring the functional status of patients with low back pain. *Arch. Phys. Med. Rehabil., 69*(12), 1044-1053.

Deyo, R. A., Gray, D. T., Kreuter, W., Mirza, S., & Martin, B. I. (2005). United States trends in lumbar fusion surgery for degenerative conditions. *Spine (Phila Pa 1976), 30*(12), 1441-1445; discussion 1446-1447.

Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One, 8*(7), e66844. doi:10.1371/journal.pone.0066844

Ekman, P., Möller, H., & Hedlund, R. (2009). Predictive factors for the outcome of fusion in adult isthmic spondylolisthesis. *Spine (Phila Pa 1976), 34*(11), 1204-1210.

Elfving, B., Andersson, T., & Grooten, W. J. (2007). Low levels of physical activity in back pain patients are associated with high levels of fear-avoidance beliefs and pain catastrophizing. *Physiother. Res. Int., 12*(1), 14-24.

Engel, G. L. (1977). The need for a new medical model: a challenge for biomedicine. *Science, 196*(4286), 129-136.

EuroQoL Group. (1990). EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy, 16*(3), 199-208.

Everhart, J. S., Best, T. M., & Flanigan, D. C. (2015). Psychological predictors of anterior cruciate ligament reconstruction outcomes: a systematic review. *Knee Surg. Sports Traumatol. Arthrosc., 23*(3), 752-762. doi:10.1007/s00167-013-2699-1

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression: Models, Methods and Applications.* Berlin: Springer Berlin Heidelberg.

Fairbank, J., Frost, H., Wilson-MacDonald, J., Yu, L.-M., Barker, K., Collins, R., & for the Spine Stabilisation Trial, G. (2005). Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *BMJ : British Medical Journal, 330*(7502), 1233-1233. doi:10.1136/bmj.38441.620417.8F

Fairbank, J. C. T., Davies, J. B., Couper, J., & O'Brien, J. P. (1980). The Oswestry low back pain disability questionnaire. *Physiotherapy, 66*(8), 271-273.

Fairbank, J. C. T., & Pynsent, P. B. (2000). The Oswestry Disability Index. *Spine (Phila Pa 1976), 25*(22), 2940-2953.

Filho, I. T., Simmonds, M. J., Protas, E. J., & Jones, S. (2002). Back pain, physical function, and estimates of aerobic capacity: what are the relationships among methods and measures? *Am. J. Phys. Med. Rehabil., 81*(12), 913-920. doi:10.1097/01.phm.0000030729.77020.2a

Fischer, D., Stewart, A. L., Bloch, D. A., Lorig, K., Laurent, D., & Holman, H. (1999). Capturing the patient's view of change as a clinical outcome measure. *JAMA, 282*(12), 1157-1162.

Food and Drug Administration. (2009). Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Retrieved from https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf

Foster, N. E., Anema, J. R., Cherkin, D., Chou, R., Cohen, S. P., Gross, D. P., Ferreira, P. H., Fritz, J. M., Koes, B. W., Peul, W., Turner, J. A., & Maher, C. G. (2018). Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet, 391*(10137), 2368-2383. doi:10.1016/s0140-6736(18)30489-6

Fritzell, P., Hagg, O., Wessberg, P., & Nordwall, A. (2001). 2001 Volvo Award Winner in Clinical Studies: Lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine (Phila Pa 1976), 26*(23), 2521-2532; discussion 2532-2524.

Fritzell, P., Hägg, O., Gerdhem, P., Abbott, A., Songsong, A., Parai, C., Thoreson, O., & Strömqvist, B. (2018). 2018 annual report follow up of spine surgery performed in Sweden in 2017. Retrieved from http://www.4s.nu/pdf/180910_Swespine_%20arsrapport_2018_eng_isbn.pdf

Froud, R., Patterson, S., Eldridge, S., Seale, C., Pincus, T., Rajendran, D., Fossum, C., & Underwood, M. (2014). A systematic review and meta-synthesis of the impact of low back pain on people's lives. *BMC Musculoskelet. Disord., 15*(1). doi:10.1186/1471-2474-15-50

Gatchel, R. J., Peng, Y. B., Peters, M. L., Fuchs, P. N., & Turk, D. C. (2007). The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychol. Bull., 133*(4), 581-624. doi:10.1037/0033-2909.133.4.581

Gatti, A. A., Stratford, P. W., Brenneman, E. C., & Maly, M. R. (2016). GT3X+ accelerometer placement affects the reliability of step-counts measured during running and pedal-revolution counts measured during bicycling. *J. Sports Sci., 34*(12), 1168-1175. doi:10.1080/02640414.2015.1096018

Gautschi, O. P., Corniola, M. V., Schaller, K., Smoll, N. R., & Stienen, M. N. (2014). The need for an objective outcome measurement in spine surgery--the timed-up-and-go test. *Spine J, 14*(10), 2521-2522. doi:10.1016/j.spinee.2014.05.004

Gautschi, O. P., Joswig, H., Corniola, M. V., Smoll, N. R., Schaller, K., Hildebrandt, G., & Stienen, M. N. (2016a). Pre- and postoperative correlation of patient-reported outcome measures with standardized Timed Up and Go (TUG) test results in lumbar degenerative disc disease. *Acta Neurochir., 158*(10), 1875-1881. doi:10.1007/s00701-016-2899-9

Gautschi, O. P., Joswig, H., Corniola, M. V., Smoll, N. R., Schaller, K., Hildebrandt, G., & Stienen, M. N. (2016b). Pre- and postoperative correlation of patient-reported outcome measures with standardized Timed Up and Go (TUG) test results in lumbar degenerative disc disease. *Acta Neurochir. (Wien.), 158*(10), 1875-1881. doi:10.1007/s00701-016-2899-9

Gautschi, O. P., Smoll, N. R., Corniola, M. V., Joswig, H., Chau, I., Hildebrandt, G., Schaller, K., & Stienen, M. N. (2016c). Validity and Reliability of a Measurement of Objective Functional Impairment in Lumbar Degenerative Disc Disease: The Timed Up and Go (TUG) Test. *Neurosurgery, 79*(2), 270-278. doi:10.1227/neu.0000000000001195

Gautschi, O. P., Stienen, M. N., Corniola, M. V., Joswig, H., Schaller, K., Hildebrandt, G., & Smoll, N. R. (2017). Assessment of the Minimum Clinically Important Difference in the Timed Up and Go Test After Surgery for Lumbar Degenerative Disc Disease. *Neurosurgery, 80*(3), 380-385. doi:10.1227/neu.0000000000001320

Getty, C. J. (1980). Lumbar spinal stenosis: the clinical spectrum and the results of operation. *The Journal of Bone and Joint Surgery. British*

volume, *62-B*(4), 481-485. doi:10.1302/0301-620X.62B4.6448861

Gilmore, S. J., McClelland, J. A., & Davidson, M. (2015). Physiotherapeutic interventions before and after surgery for degenerative lumbar conditions: a systematic review. *Physiotherapy, 101*(2), 111-118. doi:10.1016/j.physio.2014.06.007

Glassman, S. D., Anagnost, S. C., Parker, A., Burke, D., Johnson, J. R., & Dimar, J. R. (2000). The effect of cigarette smoking and smoking cessation on spinal fusion. *Spine (Phila Pa 1976), 25*(20), 2608-2615.

Glombiewski, J. A., Holzapfel, S., Riecke, J., Vlaeyen, J. W. S., de Jong, J., Lemmer, G., & Rief, W. (2018). Exposure and CBT for chronic back pain: An RCT on differential efficacy and optimal length of treatment. *J. Consult. Clin. Psychol., 86*(6), 533-545. doi:10.1037/ccp0000298

Grotle, M., Brox, J. I., & Vollestad, N. K. (2004a). Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. *Spine (Phila Pa 1976), 29*(21), E492-501.

Grotle, M., Brox, J. I., & Vollestad, N. K. (2005). Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. *Spine (Phila Pa 1976), 30*(1), 130-140.

Grotle, M., Brox, J. I., & Vøllestad, N. K. (2003). Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. *J. Rehabil. Med., 35*(5), 241-247.

Grotle, M., Vollestad, N. K., Veierod, M. B., & Brox, J. I. (2004b). Fear-avoidance beliefs and distress in relation to disability in acute and chronic low back pain. *Pain, 112*(3), 343-352. doi:10.1016/j.pain.2004.09.020

Gunzburg, R., Keller, T., Szpalski, M., Vandeputte, K., & Spratt, K. (2003). Clinical and psychofunctional measures of conservative decompression surgery for lumbar spinal stenosis: a prospective cohort study. *Eur. Spine J., 12*(2), 197-204. doi:10.1007/s00586-002-0479-6

Guralnik, J. M., Branch, L. G., Cummings, S. R., & Curb, J. D. (1989). Physical performance measures in aging research. *J. Gerontol., 44*(5), M141-146.

Guyatt, G. H., Sullivan, M. J., Thompson, P. J., Fallen, E. L., Pugsley, S. O., Taylor, D. W., & Berman, L. B. (1985). The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. *Can. Med. Assoc. J., 132*(8), 919-923.

Guyatt, G. H., Juniper, E. F., Walter, S. D., Griffith, L. E., & Goldstein, R. S. (1998). Interpreting treatment effects in randomised trials. *BMJ, 316*(7132), 690-693.

Hagstromer, M., Ainsworth, B. E., Oja, P., & Sjostrom, M. (2010). Comparison of a subjective and an objective measure of physical activity in a population sample. *J Phys Act Health, 7*(4), 541-550.

Hagströmer, M., Bergman, P., Bauman, A., & Sjöström, M. (2006). The international prevalence study (IPS): health-enhancing physical activity in Sweden. *Journal of Public Health, 14*(5), 301-308. doi:10.1007/s10389-006-0032-x

Hagströmer, M., Troiano, R. P., Sjostrom, M., & Berrigan, D. (2010). Levels and patterns of objectively assessed physical activity--a comparison between Sweden and the United States. *Am. J. Epidemiol., 171*(10), 1055-1064. doi:10.1093/aje/kwq069

Hansen, B. H., Kolle, E., Dyrstad, S. M., Holme, I., & Anderssen, S. A. (2012). Accelerometer-determined physical activity in adults and older people. *Med. Sci. Sports Exerc., 44*(2), 266-272. doi:10.1249/MSS.0b013e31822cb354

Harding, V. R., Williams, A. C., Richardson, P. H., Nicholas, M. K., Jackson, J. L., Richardson, I. H., & Pither, C. E. (1994). The development of a battery of measures for assessing physical functioning of chronic pain patients. *Pain, 58*(3), 367-375.

Hartvigsen, J., Christensen, K., & Frederiksen, H. (2003). Back pain remains a common symptom in

old age. A population-based study of 4486 Danish twins aged 70-102. *Eur. Spine J., 12*(5), 528-534. doi:10.1007/s00586-003-0542-y

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*(10137), 2356-2367. doi:10.1016/s0140-6736(18)30480-x

Hedlund, R., Johansson, C., Hagg, O., Fritzell, P., & Tullberg, T. (2016). The long-term outcome of lumbar fusion in the Swedish lumbar spine study. *Spine J, 16*(5), 579-587. doi:10.1016/j.spinee.2015.08.065

Hosmer, D. W., & Lemeshow, S. (1999). *Applied survival analysis: regression modeling of time to event data.* New York: Wiley.

Hoy, D., Bain, C., Williams, G., March, L., Brooks, P., Blyth, F., Woolf, A., Vos, T., & Buchbinder, R. (2012). A systematic review of the global prevalence of low back pain. *Arthritis Rheum., 64*(6), 2028-2037. doi:10.1002/art.34347

Hoy, D., March, L., Brooks, P., Blyth, F., Woolf, A., Bain, C., Williams, G., Smith, E., Vos, T., Barendregt, J., Murray, C., Burstein, R., & Buchbinder, R. (2014). The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann. Rheum. Dis., 73*(6), 968-974. doi:10.1136/annrheumdis-2013-204428

Hägg, O., Fritzell, P., Oden, A., & Nordwall, A. (2002). Simplifying outcome measurement: evaluation of instruments for measuring outcome after fusion surgery for chronic low back pain. *Spine (Phila Pa 1976), 27*(11), 1213-1222.

Jespersen, A. B., & Gustafsson, M. (2018). Correlation between the Oswestry Disability Index and objective measurements of walking capacity and performance in patients with lumbar spinal stenosis: a systematic literature review. *Eur. Spine J., 27*(7), 1604-1613. doi:10.1007/s00586-018-5520-5

Johansson, A. C., Ohrvik, J., & Soderlund, A. (2016). Associations among pain, disability and psychosocial factors and the predictive value of expectations on returning to work in patients who undergo lumbar disc surgery. *Eur. Spine J., 25*(1), 296-303. doi:10.1007/s00586-015-3820-6

Joswig, H., Stienen, M. N., Smoll, N. R., Corniola, M. V., Chau, I., Schaller, K., Hildebrandt, G., & Gautschi, O. P. (2017). Patients' Preference of the Timed Up and Go Test or Patient-Reported Outcome Measures Before and After Surgery for Lumbar Degenerative Disk Disease. *World Neurosurg., 99*, 26-30. doi:10.1016/j.wneu.2016.11.039

Kahraman, T., Ozcan Kahraman, B., Salik Sengul, Y., & Kalemci, O. (2016). Assessment of sit-to-stand movement in nonspecific low back pain: a comparison study for psychometric properties of field-based and laboratory-based methods. *Int. J. Rehabil. Res., 39*(2), 165-170. doi:10.1097/mrr.0000000000000164

Kalakoti, P., Missios, S., Maiti, T., Konar, S., Bir, S., Bollam, P., & Nanda, A. (2016). Inpatient Outcomes and Postoperative Complications After Primary Versus Revision Lumbar Spinal Fusion Surgeries for Degenerative Lumbar Disc Disease: A National (Nationwide) Inpatient Sample Analysis, 2002-2011. *World Neurosurg., 85*, 114-124. doi:10.1016/j.wneu.2015.08.020

Kamper, S. J., Henschke, N., Hestbaek, L., Dunn, K. M., & Williams, C. M. (2016). Musculoskeletal pain in children and adolescents. *Brazilian Journal of Physical Therapy, 20*(3), 275-284. doi:10.1590/bjpt-rbf.2014.0149

Kamper, S. J., Ostelo, R. W., Knol, D. L., Maher, C. G., de Vet, H. C., & Hancock, M. J. (2010). Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J. Clin. Epidemiol., 63*(7), 760-766.e761. doi:10.1016/j.jclinepi.2009.09.009

Kelly, L., McMillan, D., Anderson, A., Fippinger, M., Fillerup, G., & Rider, J. (2013). Validity of actigraphs uniaxial and triaxial accelerometers for

assessment of physical activity in adults in laboratory conditions. *BMC Med. Phys., 13*(1), 5.

Kemani, M., Grimby-Ekman, A., Lundgren, J., Sullivan, M., & Lundberg, M. (2018). Factor structure and internal consistency of a Swedish version of the Pain Catastrophizing Scale. *Acta Anaesthesiol. Scand.* doi:10.1111/aas.13246

Kori, K., Miller, R., & Todd, D. (1990). Kinesiophobia: a new view of chronic pain behavior. *Pain Manag, 3*, 35-43.

Kroman, S. L., Roos, E. M., Bennell, K. L., Hinman, R. S., & Dobson, F. (2014). Measurement properties of performance-based outcome measures to assess physical function in young and middle-aged people known to be at high risk of hip and/or knee osteoarthritis: a systematic review. *Osteoarthritis Cartilage, 22*(1), 26-39. doi:10.1016/j.joca.2013.10.021

Krug, S., Jordan, S., Mensink, G. B., Muters, S., Finger, J., & Lampert, T. (2013). Physical activity: results of the German Health Interview and Examination Survey for Adults (DEGS1). *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz, 56*(5-6), 765-771. doi:10.1007/s00103-012-1661-6

LaCaille, R. A., DeBerard, M. S., Masters, K. S., Colledge, A. L., & Bacon, W. (2005). Presurgical biopsychosocial factors predict multidimensional patient: outcomes of interbody cage lumbar fusion. *The Spine Journal, 5*(1), 71-78. doi:http://dx.doi.org/10.1016/j.spinee.2004.08.004

Lauridsen, H. H., Hartvigsen, J., Manniche, C., Korsholm, L., & Grunnet-Nilsson, N. (2006). Danish version of the Oswestry disability index for patients with low back pain. Part 1: Cross-cultural adaptation, reliability and validity in two different populations. *Eur. Spine J., 15*(11), 1705-1716.

Lee, C. E., Simmonds, M. J., Novy, D. M., & Jones, S. (2001). Self-reports and clinician-measured physical function among patients with low back pain: a comparison. *Arch. Phys. Med. Rehabil., 82*(2), 227-231. doi:10.1053/apmr.2001.18214

Lee, I. M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., & Katzmarzyk, P. T. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The Lancet, 380*(9838), 219-229. doi:http://dx.doi.org/10.1016/S0140-6736(12)61031-9

Lin, C. W., McAuley, J. H., Macedo, L., Barnett, D. C., Smeets, R. J., & Verbunt, J. A. (2011). Relationship between physical activity and disability in low back pain: a systematic review and meta-analysis. *Pain, 152*(3), 607-613. doi:10.1016/j.pain.2010.11.034

Lindbäck, Y., Tropp, H., Enthoven, P., Abbott, A., & Oberg, B. (2017). PREPARE: presurgery physiotherapy for patients with degenerative lumbar spine disorder: a randomized controlled trial. *Spine J.* doi:10.1016/j.spinee.2017.12.009

Lisspers, J., Nygren, A., & Soderman, E. (1997). Hospital Anxiety and Depression Scale (HAD): some psychometric data for a Swedish sample. *Acta Psychiatr. Scand., 96*(4), 281-286.

Loeser, J. D., Baron, R., Basbaum, A., Dworkin, R., Giamberardino, M. A., Kosek, E., Okifuji, A., Paice, P., & Serra, J. (2011). *Classification of Chronic Pain, Second Edition (Revised)*. Seattle: IASP Press.

Lotzke, H., Brisby, H., Gutke, A., Hägg, O., Jakobsson, M., Smeets, R., & Lundberg, M. (2018). A Person-Centered Prehabilitation Program Based on Cognitive-Behavioral Physical Therapy for Patients Scheduled for Lumbar Fusion Surgery – A Randomized Controlled Trial. *Physical Therapy Journal*. Accepted manuscript.

Lotzke, H., Jakobsson, M., Brisby, H., Gutke, A., Hägg, O., Smeets, R., den Hollander, M., Olsson, L. E., & Lundberg, M. (2016). Use of the PREPARE (PREhabilitation, Physical Activity and exeRcisE) program to improve outcomes after lumbar fusion surgery for severe low back pain: a study protocol of a person-centred randomised controlled trial. *BMC Musculoskelet. Disord., 17*(1), 349. doi:10.1186/s12891-016-1203-8

Lundberg, M., Frennered, K., Hagg, O., & Styf, J. (2011). The impact of fear-avoidance model variables on disability in patients with specific or nonspecific chronic low back pain. *Spine (Phila Pa 1976), 36*(19), 1547-1553. doi:10.1097/BRS.0b013e3181f61660

Lundberg, M., Styf, J., & Carlsson, S. G. (2004). A psychometric evaluation of the Tampa Scale for Kinesiophobia - From a physiotherapeutic perspective. *Physiother. Theory Pract., 20*(2), 121-133.

Luzak, A., Heier, M., Thorand, B., Laxy, M., Nowak, D., Peters, A., & Schulz, H. (2017). Physical activity levels, duration pattern and adherence to WHO recommendations in German adults. *PLoS One, 12*(2), e0172503. doi:10.1371/journal.pone.0172503

Madera, M., Brady, J., Deily, S., McGinty, T., Moroz, L., Singh, D., Tipton, G., & Truumees, E. (2017). The role of physical therapy and rehabilitation after lumbar fusion surgery for degenerative disease: a systematic review. *J. Neurosurg. Spine, 26*(6), 694-704. doi:10.3171/2016.10.Spine16627

Magnussen, L., Strand, L. I., & Lygren, H. (2004). Reliability and validity of the back performance scale: observing activity limitation in patients with back pain. *Spine (Phila Pa 1976), 29*(8), 903-907.

Maher, C., Underwood, M., & Buchbinder, R. (2017). Non-specific low back pain. *The Lancet, 389*(10070), 736-747. doi:https://doi.org/10.1016/S0140-6736(16)30970-9

Mannion, A. F., Brox, J. I., & Fairbank, J. C. (2016). Consensus at last! Long-term results of all randomized controlled trials show that fusion is no better than non-operative care in improving pain and disability in chronic low back pain. *Spine J, 16*(5), 588-590. doi:10.1016/j.spinee.2015.12.001

Mannion, A. F., & Elfering, A. (2006). Predictors of surgical outcome and their assessment. *Eur. Spine J., 15*(SUPPL. 1), S93-S108.

Mayer, T. G., Barnes, D., Kishino, N. D., Nichols, G., Gatchel, R. J., Mayer, H., & Mooney, V.

(1988). Progressive isoinertial lifting evaluation. I. A standardized protocol and normative database. *Spine (Phila Pa 1976), 13*(9), 993-997.

Mobbs, R. J., Phan, K., Maharaj, M., & Rao, P. J. (2016). Physical Activity Measured with Accelerometer and Self-Rated Disability in Lumbar Spine Surgery: A Prospective Study. *Global Spine Journal, 6*(5), 459-464. doi:10.1055/s-0035-1565259

Modic, M. T., & Ross, J. S. (2007). Lumbar Degenerative Disk Disease. *Radiology, 245*(1), 43-61. doi:10.1148/radiol.2451051706

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010a). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J. Clin. Epidemiol., 63*(7), 737-745. doi:10.1016/j.jclinepi.2010.02.006

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010b). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual. Life Res., 19*(4), 539-549. doi:10.1007/s11136-010-9606-8

Montero-Odasso, M., Schapira, M., Soriano, E. R., Varela, M., Kaplan, R., Camera, L. A., & Mayorga, L. M. (2005). Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older. *J. Gerontol. A Biol. Sci. Med. Sci., 60*(10), 1304-1309.

Norden, J., Smuck, M., Sinha, A., Hu, R., & Tomkins-Lane, C. (2017). Objective measurement of free-living physical activity (performance) in lumbar spinal stenosis: are physical activity guidelines being met? *Spine J, 17*(1), 26-33. doi:10.1016/j.spinee.2016.10.016

Ocarino, J., Gonçalves, G., Vaz, D., Cabral, A., Porto, J., & Silva, M. (2009). Correlation between a functional performance questionnaire and physical performance tests among patients with low

back pain. *Brazilian Journal of Physical Therapy, 13*(4), 343-349.

Odebiyi, D. O., Kujero, S. O., & Lawal, T. A. (2007). Relationship Between Spinal Mobility, Physical Performance, Pain Intensity and Functional Disability In Patients With Chronic Low Back Pain. *Nigerian Journal of Medical Rehabilitation, 11*(2), 49-54.

Ostir, G. V., Kuo, Y. F., Berges, I. M., Markides, K. S., & Ottenbacher, K. J. (2007). Measures of lower body function and risk of mortality over 7 years of follow-up. *Am. J. Epidemiol., 166*(5), 599-605. doi:10.1093/aje/kwm121

Ozemek, C., Kirschner, M. M., Wilkerson, B. S., Byun, W., & Kaminsky, L. A. (2014). Inter-monitor reliability of the GT3X+ accelerometer at hip, wrist and ankle sites during activities of daily living. *Physiol. Meas., 35*(2), 129-138. doi:10.1088/0967-3334/35/2/129

Pekkanen, L., Kautiainen, H., Ylinen, J., Salo, P., & Hakkinen, A. (2011). Reliability and validity study of the Finnish version 2.0 of the oswestry disability index. *Spine (Phila Pa 1976), 36*(4), 332-338. doi:10.1097/BRS.0b013e3181cdd702

Peters, M. L., Vlaeyen, J. W., & Weber, W. E. (2005). The joint contribution of physical pathology, pain-related fear and catastrophizing to chronic back pain disability. *Pain, 113*(1-2), 45-50. doi:10.1016/j.pain.2004.09.033

Phillips, F. M., Slosar, P. J., Youssef, J. A., Andersson, G., & Papatheofanis, F. (2013). Lumbar Spine Fusion for Chronic Low Back Pain Due to Degenerative Disc Disease: A Systematic Review. *Spine (Phila Pa 1976), 38*(7), E409-E422. doi:10.1097/BRS.0b013e3182877f11

Pincus, T., Vlaeyen, J. W., Kendall, N. A., Von Korff, M. R., Kalauokalani, D. A., & Reis, S. (2002). Cognitive-behavioral therapy and psychosocial factors in low back pain: directions for the future. *Spine (Phila Pa 1976), 27*(5), E133-138.

Pratt, R. K., Fairbank, J. C., & Virr, A. (2002). The reliability of the Shuttle Walking Test, the Swiss Spinal Stenosis Questionnaire, the Oxford Spinal Stenosis Score, and the Oswestry Disability Index in the assessment of patients with lumbar spinal stenosis. *Spine (Phila Pa 1976), 27*(1), 84-91.

Price, D. D., McGrath, P. A., Rafii, A., & Buckingham, B. (1983). The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain, 17*(1), 45-56.

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int. J. Behav. Nutr. Phys. Act., 5*, 56. doi:10.1186/1479-5868-5-56

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual. Life Res., 27*(5), 1147-1157. doi:10.1007/s11136-018-1798-3

Prinsen, C. A. C., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., Williamson, P. R., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. *Trials, 17*(1), 449. doi:10.1186/s13063-016-1555-2

Qaseem, A., Wilt, T. J., McLean, R. M., & Forciea, M. A. (2017). Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain: A Clinical Practice Guideline From the American College of Physicians. *Ann. Intern. Med., 166*(7), 514-530. doi:10.7326/m16-2367

Rainville, J., Childs, L. A., Pena, E. B., Suri, P., Limke, J. C., Jouve, C., & Hunter, D. J. (2012). Quantification of walking ability in subjects with neurogenic claudication from lumbar spinal stenosis--a comparative study. *Spine J, 12*(2), 101-109. doi:10.1016/j.spinee.2011.12.006

Rajaee, S. S., Bae, H. W., Kanim, L. E., & Delamarter, R. B. (2012). Spinal fusion in the United States: analysis of trends from 1998 to 2008. *Spine (Phila Pa 1976), 37*(1), 67-76. doi:10.1097/BRS.0b013e31820cccfb

Rhodes, R. E., & Quinlan, A. (2015). Predictors of physical activity change among adults using observational designs. *Sports Med., 45*(3), 423-441. doi:10.1007/s40279-014-0275-6

Roland, M., & Morris, R. (1983). A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976), 8*(2), 141-144.

Rolving, N., Obling, K. H., Christensen, F. B., & Fonager, K. (2013). Physical activity level, leisure activities and related quality of life 1 year after lumbar decompression or total hip arthroplasty. *Eur. Spine J., 22*(4), 802-808. doi:10.1007/s00586-012-2535-1

Rydwik, E., Hovmoller, F., & Bostrom, C. (2014). Aspects of reliability and validity of the Swedish version of the Self-Efficacy for Exercise Scale for older people. *Physiother. Theory Pract., 30*(2), 131-137. doi:10.3109/09593985.2013.838614

Sallis, J. F., Owen, N., & Fisher, E. B. (2008). Ecological models of health behavior. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health behavior and health education: theory, research, and practice.* Hoboken: John Wiley & Sons, Inc.

Scott, W., & McCracken, L. M. (2015). Patients' Impression of Change Following Treatment for Chronic Pain: Global, Specific, a Single Dimension, or Many? *The Journal of Pain, 16*(6), 518-526. doi:https://doi.org/10.1016/j.jpain.2015.02.007

Seel, R. T., Steyerberg, E. W., Malec, J. F., Sherer, M., & Macciocchi, S. N. (2012). Developing and evaluating prediction models in rehabilitation populations. *Arch. Phys. Med. Rehabil., 93*(8 Suppl), S138-153. doi:10.1016/j.apmr.2012.04.021

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.*

Shiri, R., Karppinen, J., Leino-Arjas, P., Solovieva, S., & Viikari-Juntura, E. (2010). The association between obesity and low back pain: a meta-analysis. *Am. J. Epidemiol., 171*(2), 135-154. doi:10.1093/aje/kwp356

Simmonds, M. J., Olson, S. L., Jones, S., Hussein, T., Lee, C. E., Novy, D., & Radwan, H. (1998). Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. *Spine (Phila Pa 1976), 23*(22), 2412-2421.

Slootmaker, S., Schuit, A., Chinapaw, M., Seidell, J., & van Mechelen, W. (2009). Disagreement in physical activity assessed by accelerometer and self-report in subgroups of age, gender, education and weight status. *International Journal of Behavioral Nutrition and Physical Activity, 6*(1), 17.

Smeets, R., Hijdra, H. J. M., Kester, A. D. M., Hitters, M. W. G. C., & Knottnerus, J. A. (2006). The usability of six physical performance tasks in a rehabilitation population with chronic low back pain. *Clin. Rehabil., 20*(11), 989-998.

Smeets, R., van Geel, A. C., Kester, A. D., & Knottnerus, J. A. (2007). Physical capacity tasks in chronic low back pain: what is the contributing role of cardiovascular capacity, pain and psychological factors? *Disabil. Rehabil., 29*(7), 577-586. doi:10.1080/09638280600925829

Smuck, M., Muaremi, A., Zheng, P., Norden, J., Sinha, A., Hu, R., & Tomkins-Lane, C. (2018). Objective measurement of function following lumbar spinal stenosis decompression reveals improved functional capacity with stagnant real-life physical activity. *Spine J, 18*(1), 15-21. doi:10.1016/j.spinee.2017.08.262

Soer, R., Poels, B. J., Geertzen, J. H., & Reneman, M. F. (2006). A comparison of two lifting assessment approaches in patients with chronic low back pain. *J Occup Rehabil, 16*(4), 639-646. doi:10.1007/s10926-006-9055-y

Somerson, J. S., Bartush, K. C., Shroff, J. B., Bhandari, M., & Zelle, B. A. (2016). Loss to follow-up in orthopaedic clinical trials: a systematic review. *Int. Orthop., 40*(11), 2213-2219. doi:10.1007/s00264-016-3212-5

Soriano, J. C., Sendino, M. R., Fabregate, M. F., Cimarra, I. D., Martínez, P. U., & Deglané, R. M. (2010). Predictors of outcome after decom-

pressive lumbar surgery and instrumented posterolateral fusion. *Eur. Spine J., 19*(11), 1841-1848. doi:10.1007/s00586-010-1284-2

Staartjes, V. E., & Schroder, M. L. (2018). The five-repetition sit-to-stand test: evaluation of a simple and objective tool for the assessment of degenerative pathologies of the lumbar spine. *J. Neurosurg. Spine*, 1-8. doi:10.3171/2018.2.Spine171416

Steyerberg, E. W. (2009). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.* New York: Springer.

Strand, L. I., Anderson, B., Lygren, H., Skouen, J. S., Ostelo, R., & Magnussen, L. H. (2011). Responsiveness to change of 10 physical tests used for patients with back pain. *Phys. Ther., 91*(3), 404-415. doi:10.2522/ptj.20100016

Strand, L. I., Moe-Nilssen, R., & Ljunggren, A. E. (2002). Back Performance Scale for the assessment of mobility-related activities in people with back pain. *Phys. Ther., 82*(12), 1213-1223.

Strath, S. J., Kaminsky, L. A., Ainsworth, B. E., Ekelund, U., Freedson, P. S., Gary, R. A., Richardson, C. R., Smith, D. T., & Swartz, A. M. (2013). Guide to the assessment of physical activity: Clinical and research applications: a scientific statement from the American Heart Association. *Circulation, 128*(20), 2259-2279. doi:10.1161/01.cir.0000435708.67487.da

Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: a practical guide to their development and use.* Oxford: Oxford University Press.

Strömqvist, B., Fritzell, P., Hagg, O., Jonsson, B., & Sanden, B. (2013). Swespine, the Swedish spine register the – 2012 report. *Eur. Spine J., 22*(4), 953-974. doi:10.1007/s00586-013-2758-9

Sullivan, M. J., Adams, H., Rhodenizer, T., & Stanish, W. D. (2006). A psychosocial risk factor-targeted intervention for the prevention of chronic pain and disability following whiplash injury. *Phys. Ther., 86*(1), 8-18.

Sullivan, M. J., Thorn, B., Haythornthwaite, J. A., Keefe, F., Martin, M., Bradley, L. A., & Lefebvre, J. C. (2001). Theoretical perspectives on the relation between catastrophizing and pain. *Clin. J. Pain, 17*(1), 52-64.

Swartz, A. M., Rote, A. E., Cho, Y. I., Welch, W. A., & Strath, S. J. (2014). Responsiveness of motion sensors to detect change in sedentary and physical activity behaviour. *Br. J. Sports Med., 48*(13), 1043-1047. doi:10.1136/bjsports-2014-093520

Tabue-Teguo, M., Le Goff, M., Avila-Funes, J. A., Frison, E., Helmer, C., Feart, C., Amieva, H., & Dartigues, J. F. (2015). Walking and psychomotor speed in the elderly: concordance, correlates and prediction of death. *J. Nutr. Health Aging, 19*(4), 468-473. doi:10.1007/s12603-014-0560-y

Taylor, S., Frost, H., Taylor, A., & Barker, K. (2001). Reliability and responsiveness of the shuttle walking test in patients with chronic low back pain. *Physiother. Res. Int., 6*(3), 170-178.

Teixeira Da Cunha-Filho, I., Lima, F. C., Guimarães, F. R., & Leite, H. R. (2010). Use of physical performance tests in a group of Brazilian Portuguese-speaking individuals with low back pain. *Physiother. Theory Pract., 26*(1), 49-55.

Terwee, C., Mokkink, L., Knol, D., Ostelo, R. J. G., Bouter, L., & de Vet, H. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual. Life Res., 21*(4), 651-657. doi:10.1007/s11136-011-9960-1

Terwee, C. B., Roorda, L. D., Dekker, J., Bierma-Zeinstra, S. M., Peat, G., Jordan, K. P., Croft, P., & de Vet, H. C. (2010). Mind the MIC: large variation among populations and methods. *J. Clin. Epidemiol., 63*(5), 524-534. doi:10.1016/j.jclinepi.2009.08.010

The Health and Social Care Information Centre. (2011). HES online. Primary. Diagnosis: 3 Character Tables. Retrieved from https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics

Tomkins, C. C., Battie, M. C., Rogers, T., Jiang, H., & Petersen, S. (2009). A criterion measure of walking capacity in lumbar spinal stenosis and its comparison with a treadmill protocol. *Spine (Phila Pa 1976), 34*(22), 2444-2449. doi:10.1097/BRS.0b013e3181b03fc8

Trief, P. M., Grant, W., & Fredrickson, B. (2000). A prospective study of psychological predictors of lumbar surgery outcome. *Spine (Phila Pa 1976), 25*(20), 2616-2621.

Trief, P. M., Ploutz-Snyder, R., & Fredrickson, B. E. (2006). Emotional health predicts pain and function after fusion: A prospective multicenter study. *Spine (Phila Pa 1976), 31*(7), 823-830.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Med. Sci. Sports Exerc., 40*(1), 181-188. doi:10.1249/mss.0b013e31815a51b3

Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Med. Sci. Sports Exerc., 37*(11 Suppl), S531-543.

Tudor-Locke, C., Craig, C. L., Thyfault, J. P., & Spence, J. C. (2013). A step-defined sedentary lifestyle index: <5000 steps/day. *Appl. Physiol. Nutr. Metab., 38*(2), 100-114. doi:10.1139/apnm-2012-0235

Turner, J. A., Jensen, M. P., & Romano, J. M. (2000). Do beliefs, coping, and catastrophizing independently predict functioning in patients with chronic pain? *Pain, 85*(1-2), 115-125.

Waddell, G. (1992). Biopsychosocial analysis of low back pain. *Baillieres Clin. Rheumatol., 6*(3), 523-557. doi:https://doi.org/10.1016/S0950-3579(05)80126-8

Valderas, J. M., Kotzeva, A., Espallargues, M., Guyatt, G., Ferrans, C. E., Halyard, M. Y., Revicki, D. A., Symonds, T., Parada, A., & Alonso, J. (2008). The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual. Life Res., 17*(2), 179-193. doi:10.1007/s11136-007-9295-0

van den Akker-Scheek, I., Stevens, M., Groothoff, J. W., Bulstra, S. K., & Zijlstra, W. (2007). Preoperative or postoperative self-efficacy: which is a better predictor of outcome after total hip or knee arthroplasty? *Patient Educ. Couns., 66*(1), 92-99. doi:10.1016/j.pec.2006.10.012

van der Ploeg, H. P., van der Beek, A. J., van der Woude, L. H., & van Mechelen, W. (2004). Physical activity for people with a disability: a conceptual model. *Sports Med., 34*(10), 639-649.

van Susante, J., van De Schaaf, D., & Pavlov, P. (1998). Psychological distress deteriorates the subjective outcome of lumbosacral fusion. A prospective study. *Acta Orthop. Belg., 64*(4), 371-377.

van Tulder, M., Becker, A., Bekkering, T., Breen, A., del Real, M. T., Hutchinson, A., Koes, B., Laerum, E., & Malmivaara, A. (2006). Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. *Eur. Spine J., 15 Suppl 2*, S169-191. doi:10.1007/s00586-006-1071-2

van Weering, M. G. H., Vollenbroek-Hutten, M. M. R., Tönis, T. M., & Hermens, H. J. (2009). Daily physical activities in chronic lower back pain patients assessed with accelerometry. *European Journal of Pain, 13*(6), 649-654. doi:10.1016/j.ejpain.2008.07.005

Wand, B. M., Chiffelle, L. A., O'Connell, N. E., McAuley, J. H., & Desouza, L. H. (2010). Self-reported assessment of disability and performance-based assessment of disability are influenced by different patient characteristics in acute low back pain. *Eur. Spine J., 19*(4), 633-640. doi:10.1007/s00586-009-1180-9

Ward, M. M., Guthrie, L. C., & Alba, M. (2015). Domain-specific transition questions demonstrated higher validity than global transition questions as anchors for clinically important improvement. *J. Clin. Epidemiol., 68*(6), 655-661. doi:10.1016/j.jclinepi.2015.01.028

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med. Care, 30*(6), 473-483.

Weiss, A. J., Elixhauser, A., & Andrews, R. M. (2014). Characteristics of Operating Room Procedures in U.S. Hospitals, 2011: Statistical Brief #170. In *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs.* Rockville (MD): Agency for Healthcare Research and Quality (US).

Welk, G. (2002). *Physical activity assessments for health-related research.* Champaign: Human Kinetics.

Wen, C. P., Wai, J. P. M., Tsai, M. K., Yang, Y. C., Cheng, T. Y. D., Lee, M.-C., Chan, H. T., Tsao, C. K., Tsai, S. P., & Wu, X. (2011). Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. *The Lancet, 378*(9798), 1244-1253. doi:10.1016/S0140-6736(11)60749-6

Verbunt, J. A., Sieben, J. M., Seelen, H. A. M., Vlaeyen, J. W. S., Bousema, E. J., van der Heijden, G. J., & Knottnerus, J. A. (2005). Decline in physical activity, disability and pain-related fear in subacute low back pain. *European Journal of Pain, 9*(4), 417-417. doi:10.1016/j.ejpain.2004.09.011

Verbunt, J. A., Smeets, R. J., & Wittink, H. M. (2010). Cause or effect? Deconditioning and chronic low back pain. *Pain, 149*(3), 428-430. doi:10.1016/j.pain.2010.01.020

Verbunt, J. A., Westerterp, K. R., van der Heijden, G. J., Seelen, H. A., Vlaeyen, J. W., & Knottnerus, J. A. (2001). Physical activity in daily life in patients with chronic low back pain. *Arch. Phys. Med. Rehabil., 82*(6), 726-730. doi:10.1053/apmr.2001.23182

Vergnes, J. N., Marchal-Sixou, C., Nabet, C., Maret, D., & Hamel, O. (2010). Ethics in systematic reviews. *J. Med. Ethics, 36*(12), 771-774. doi:10.1136/jme.2010.039941

Whitehurst, M., Brown, L. E., Eidelson, S. G., & D'Angelo, A. (2001). Functional mobility performance in an elderly population with lumbar spinal stenosis. *Arch. Phys. Med. Rehabil., 82*(4), 464-467. doi:10.1053/apmr.2001.20828

Wilhelm, M., Reiman, M., Goode, A., Richardson, W., Brown, C., Vaughn, D., & Cook, C. (2017). Psychological Predictors of Outcomes with Lumbar Spinal Fusion: A Systematic Literature Review. *Physiother. Res. Int., 22*(2). doi:10.1002/pri.1648

Willems, P., de Bie, R., Öner, C., Castelein, R., & de Kleuver, M. (2011). Clinical decision making in spinal fusion for chronic low back pain. Results of a nationwide survey among spine surgeons. *BMJ Open, 1*(2). doi:10.1136/bmjopen-2011-000391

Wittink, H. (2005). Functional capacity testing in patients with chronic pain. *Clin. J. Pain, 21*(3), 197-199.

Vlaeyen, J. W., & Linton, S. J. (2012). Fear-avoidance model of chronic musculoskeletal pain: 12 years on. *Pain, 153*(6), 1144-1147. doi:10.1016/j.pain.2011.12.009

Vlaeyen, J. W. S., Kole-Snijders, A. M. J., Boeren, R. G. B., & van Eek, H. (1995). Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. *Pain, 62*(3), 363-372.

Woby, S. R., Urmston, M., & Watson, P. J. (2007). Self-efficacy mediates the relation between pain-related fear and outcome in chronic low back pain patients. *Eur. J. Pain, 11*(7), 711-718. doi:10.1016/j.ejpain.2006.10.009

Woertgen, C., Rothoerl, R. D., Breme, K., Altmeppen, J., Holzschuh, M., & Brawanski, A. (1999). Variability of outcome after lumbar disc surgery. *Spine (Phila Pa 1976), 24*(8), 807-811.

Von Korff, M., Crane, P., Lane, M., Miglioretti, D. L., Simon, G., Saunders, K., Stang, P., Brandenburg, N., & Kessler, R. (2005). Chronic spinal pain and physical-mental comorbidity in the United States: results from the national comorbidity survey replication. *Pain, 113*(3), 331-339. doi:10.1016/j.pain.2004.11.010

World Health Organization. (2001). *The International Classification of Functioning, Disability and Health (ICF). 2nd ed.* Geneva, Switzerland: World Health Organization.

World Health Organization. (2009a). Global health risks – Mortality and burden of disease attributable to selected major risks. Retrieved from https://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf

World Health Organization. (2009b). Global recommendations on physical activity for health. Retrieved from http://www.who.int/dietphysicalactivity/publications/recommendations18_64yearsold/en/

Wylde, V., Dixon, S., & Blom, A. W. (2012). The role of preoperative self-efficacy in predicting outcome after total knee replacement. *Musculoskeletal Care, 10*(2), 110-118. doi:10.1002/msc.1008

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatr. Scand., 67*(6), 361-370. doi:10.1111/j.1600-0447.1983.tb09716.x