Summary
Clinical research in low back pain: what and how should we measure?

Low back pain (LBP) represents the leading global cause of disability, according to the Global Burden of Disease study of 2010, 2015 and 2016. LBP afflicts more than 50% of the population at least once in the lifetime and its point prevalence is estimated to be around 10%. Costs associated with LBP represent a substantial burden to health care systems and to society as a whole. Since a specific cause or pathology cannot be found in the large majority (about 90%) of patients with LBP, these patients are labelled as having non-specific LBP. Finding effective health interventions for patients with non-specific LBP is a key priority for researchers, healthcare professionals and policy makers.

Randomized clinical trials represent the most robust study design to evaluate the efficacy or effectiveness of interventions. However, “clinical trials are only as credible as their endpoints”, meaning that fit-for-purpose outcome measurement instruments should be selected to evaluate interventions’ effectiveness. First, outcome instruments used in a trial should measure health aspects that are relevant for all key stakeholders, particularly patients. Second, they should be valid, reliable and responsive reflections of these aspects. Selecting instruments for a clinical trial on a specific health condition can be facilitated by the use of core outcome sets (COS).

A COS is an internationally agreed minimum set of outcomes that should be measured and reported in every clinical trial on a single health condition, with the main goal of facilitating comparison of results of different trials. The development of a COS consists of two main phases in which consensus should be achieved, first, on ‘what’ to measure (i.e. core domain set), and, second, on ‘how’ to measure (i.e. core outcome measurement set) (Chapter 1). A COS for LBP was already proposed 20 years ago, but some methodological limitations (e.g. no patient involvement) led an international group of researchers and clinicians to decide to perform an update of that COS. The update of the existing COS for LBP was the focus of this dissertation.

A core domain set for low back pain

An international steering group of researchers, clinicians and patient consumers was formed to work alongside this project (Chapter 2 and 3). To determine a core domain set to be used in clinical trials in patients with non-specific LBP, a list of potential core outcome domains with definitions was drafted using the OMERACT Filter 2.0 framework. This list was presented in a 3-round Delphi survey involving researchers, clinicians and patients, during which consensus was sought. The
Delphi study led to agreement on recommending four core outcome domains for clinical trials in non-specific LBP: physical functioning, pain intensity, health-related quality of life, and number of deaths.

**Measurement properties of instruments in low back pain**

Sets of commonly used and/or previously recommended patient-reported outcome measures (PROMs) were selected as potential core outcome measurement instruments for the COS. Their measurement properties (i.e. validity, reliability and responsiveness) were summarized in four systematic reviews focusing on studies in patients with non-specific LBP and using the recently updated COSMIN methodology. Their feasibility (e.g. length, costs) was assessed by the steering group, to facilitate their future implementation in the COS.

The first review (Chapter 4) focused solely on head-to-head comparison studies of two PROMs for physical functioning: the 24-item Roland Morris Disability Questionnaire (RMDQ-24) and the Oswestry Disability Index version 2.1a (ODI 2.1a). The ODI 2.1a displayed better reliability but the RMDQ-24 displayed better construct validity, leading to the conclusion that there is no evidence clearly suggesting that one of the two has superior measurement properties. A second systematic review (Chapter 5) focused on content and structural validity of 17 physical functioning PROMs, highlighting the lack of studies assessing their content validity in patients with LBP. Meanwhile, based on high quality evidence, structural validity was found to be problematic for three of these instruments, among which the RMDQ-24.

A third systematic review (Chapter 6) evaluated all measurement properties of Visual Analogue Scale, Numeric Rating Scale (NRS), and Pain Severity subscale of the Brief Pain Inventory to measure pain intensity in patients with LBP. This review exhibited the overall lack of evidence on the majority of measurement properties but high quality evidence was found for an insufficient (i.e. too large) measurement error of the NRS. A fourth review (Chapter 7) assessed five generic PROMs for health-related quality of life, highlighting the poor construct validity of the Short Form 36 and EuroQol-5D scores. This review also showed the overall lack of (high quality) evidence on most measurement properties of these PROMs in LBP.

**A core outcome measurement set for low back pain**

A 2-round Delphi survey (Chapter 8) involving researchers, clinicians and patients was run to seek consensus on which PROMs to endorse as core outcome measurement instruments for LBP clinical trials. A summary of their measurement properties and feasibility was presented to the Delphi participants. Consensus was achieved on one instrument for physical functioning (i.e. ODI 2.1a) and one
for pain intensity (i.e. NRS). For pain intensity, consensus was also reached on recommending a specific NRS version referring to average low back pain intensity over the last week. For health-related quality of life, consensus could not be achieved on any instrument, with the Short Form 12 (SF-12) being the only to approach the threshold. Various participants expressed their uneasiness about recommending only core instruments with a fee.

Recommendations on core outcome measurement instruments were formulated by the steering group that opted for recommending at least one free-of-charge instrument per domain. ODI 2.1a and RMDQ-24 were recommended for physical functioning, NRS for pain intensity, and SF-12 and 10-item PROMIS Global Health (PROMIS-GH-10) short form for health-related quality of life.

**Assessment of PROMs measurement properties**

The systematic review on the measurement properties of PROMs for LBP displayed the lack of head-to-head comparison studies, which would give more clear indication on whether an instrument is better than others. In this dissertation, a head-to-head example study ({Chapter 9}) compared the psychometric performance of RMDQ-24 with that of PROMIS Physical Function short forms. This study showed that the 6-, 8- and 10-item PROMIS Physical Function short forms exhibited a slightly better performance than the RMDQ-24 and the 4- and 20-item PROMIS forms in patients with chronic LBP.

Other two examples of clinimetric studies were presented in this dissertation. They both focused on the Pain Self-Efficacy Questionnaire (PSEQ). ‘Pain self-efficacy’ refers to the degree to which an individual can perform his/her daily activities, despite the presence of pain. This construct was demonstrated in various studies to be an important prognostic factor and treatment mediator in patients with LBP. The first study ({Chapter 10}) showed the cross-cultural adaptation (into Italian), validity and reliability assessment of the PSEQ, highlighting the importance of performing such process when a PROM is used in a new language. The second study ({Chapter 11}) was a head-to-head comparison of responsiveness and interpretability of the PSEQ and its short forms. It was shown that the PSEQ short forms are responsive like the original PSEQ. The results of the two studies together show that the PSEQ is a valid, reliable and responsive instrument in patients with LBP, although content validity was not assessed.

{Chapter 12} discussed the results of the core set with respect to different viewpoints around the selection of PROMs, specifically referring to their future implementation in clinical trials. Additionally, it is presented the available evidence on how to best interpret results for these instruments in clinical trials in low back pain.
Conclusions
A core outcome measurement set was established for clinical trials in patients with non-specific LBP and it should be used in every clinical trial on this condition to facilitate comparability of findings. The next step in the development of this core set is to determine how to interpret results for the core instruments in trials. Nevertheless, it should be noted that this core set does not provide final answers on core outcome domains and measurement instruments for LBP. In fact, the development of a core set is an iterative process that should be updated as soon as new (high quality) evidence emerges on relevant domains and measurement properties of the instruments.