Chapter 12

Summary & General discussion
The ultimate aim of the research presented in this thesis is to contribute to the body of knowledge on outcomes of interventions for chronic low back pain (CLBP) and to identify outcome-based subgroups of patients having different profiles. Based on work included in this thesis, an article was published in a provincial newspaper in 2014 highlighting the development of a tool to triage the right patient to the right medical specialist, based on subgroups with different patient profiles, i.e. the Nijmegen Decision Tool for CLBP. The headline stated ‘A questionnaire helps in low back pain’ [1]. Whilst we think we have taken an important step forward in terms of the knowledge gained and identifying pre-treatment profiles related to outcomes of interventions, a questionnaire as such is only a small piece of the puzzle towards solving the tremendous worldwide burden of low back pain.

Despite decades of research, honed expertise, and improved quality of clinical trials, the evidence regarding the effectiveness of the treatments offered to patients with CLBP is still inconsistent [2-8], rarely shows more than a small to moderate overall benefit [2,7,9-12], and demonstrates a lack of long-term efficacy in changing the prognostic paths [10]. Most patients recover, but in around 20% the complaints persist for more than three months, resulting in disabling CLBP. These patients bear the greatest proportion of the disease burden. One reason for the overall disappointing outcomes is that the CLBP population is heterogeneous because the condition lacks diagnostic clarity; as a consequence, there exists a plethora of invasive and non-invasive interventions in secondary or tertiary healthcare for the same symptom. To reduce the global societal burden of CLBP it is crucial to improve treatment outcomes and to reduce the related costs, to improve the value of delivered healthcare. To achieve this it is essential to know which outcomes are relevant to both patients and medical specialists and to know which patients will benefit from spine surgery or from non-surgical treatments.

In the General introduction of this thesis the aims are described using three separate, though related themes. In the first theme a non-surgical combined physical and psychological (CPP) programme was evaluated and insight is given as to who could benefit from this programme. In the second theme the focus is on the methodology used for outcomes assessment in the evaluation of clinical practice and research of degenerative lumbar spine disorders. In the third theme the development of a clinical decision tool for patient triage to a surgical or a non-surgical medical specialist is described. In this Summary & General discussion the overall research questions are answered by summarising and discussing the main findings per theme. Considerations are described that warrant further exploration in relationship to the methodology used, and some implications and recommendations for clinical practice and for future research are presented. The discussion ends with concluding remarks.
Theme A: Introduction of a combined physical and psychological programme

Research questions
1. Does the novel CPP programme for CLBP improve patient outcomes and reduce healthcare consumption?
2. Is it possible to identify a subgroup of patients that benefits most from the novel CPP programme so that selection criteria can be optimised?

Summary of main findings
In the one- and two-year follow-up cohort studies presented in Chapter 2 and 3, selected and motivated patients with longstanding CLBP participated in an intensive multidisciplinary CPP programme. Participating patients were moderately to severely disabled, which is comparable to patients being treated in secondary spine care. They learned to manage their CLBP, and their daily functioning and quality of life improved meaningfully. The magnitude of the improvement is comparable to that achieved with spinal surgery (Standardised Morbidity Ratio [SMR] 98%) and better than that achieved with less intensive rehabilitation programmes (SMR 136%) (Chapter 2). At the two-year follow-up of this cohort, the significant and clinically relevant improvements in functional ability, pain and health-related quality of life achieved at the one-year follow-up assessment were maintained. Above all, most of the participants were employed and the results indicate that the use of both pain medication and healthcare decreased substantially (Chapter 3). In Chapter 4, pre-treatment indicators of a successful treatment outcome were identified. Successful treatment outcome was defined as a one-year follow-up functional disability (disability) score falling to values seen in healthy populations (Oswestry Disability Index, ODI ≤22). Patients who are employed at pre-treatment [OR 3.61 (95% CI 1.80–7.26) and who are mild to moderately disabled at the start of a CPP programme [OR 0.94 (95% CI 0.92–0.97)] are most likely to benefit from this programme ($R^2=22%$; 67% correctly classified). No interaction effects between pre-treatment characteristics were found and, to our surprise, no predictive value was found for psychological distress.

As continuous outcomes monitoring is part of the CPP programme, this gave us the opportunity to further substantiate the ‘pilot’ results presented in Chapter 2 with one-year follow-up results of a large cohort (n=848). The results of this recent study showed that these patients had similarly good results: patients improved during the programme, showed further improvement at the one-year follow-up, and half of the patients (51%, n=433) improved such that their functional status was comparable to that of the healthy population [13].

Discussion
In a recent systematic review, moderate-quality evidence for moderate effects of multidisciplinary bio-psychosocial treatments was found compared to usual care [12]. These treatments are recommended for CLBP patients [14-16], but not often implemented.

In this thesis the clinical relevance of the CPP programme has been demonstrated: patients improved meaningfully, healthcare consumption was reduced, and a relevant treatment effect was found. We found large effect sizes for functional ability at the one-year follow-up (Chapter 2) and the two-year follow-up (Chapter 3), meaning that the programme shows relevant effects. The two-year follow-up findings presented in Chapter 3 were further substantiated.
Summary and General discussion

by the results of a recently performed long-term follow-up post-marketing surveillance study involving 277 ex-participants (mean follow-up of 6.5 years [range: 5.5–7.5]; response 85%; no baseline differences between responders and non-responders to the survey); positive results were maintained after 6.5 years on average, and 80% of the participants were satisfied with the treatment results [17].

We studied the intervention as an integral programme. The programme uses a wide range of techniques based on cognitive behavioural principles. As yet, the working mechanisms are unknown, so it is unclear which techniques or parts of the intervention are the effective elements. In Chapter 4 we speculate that several aspects contribute to the success of this programme, such as the programme’s structure (a. the intensity or dose of the programme), the programme entrance criteria (b. motivation to change behaviour), and the content of the programme (c. improvement of dysfunctional cognitive behavioural factors).

a) **Intensity of the programme.** Although conflicting evidence exists concerning stability over time [18,19], it is assumed that dysfunctional behavioural cognitions in patients with persistent pain of long duration are resistant to change [20]. In a systematic review, Guzman et al. recommend 100 hours or more of intensive multidisciplinary rehabilitation including cognitive behavioural interventions to improve functionality [21]. The CPP programme studied in this thesis follows this recommendation in a ‘pressure-cooker’ structure. This ‘pressure-cooker’ structure might moderate the dysfunctional behavioural cognitions. Patients with longstanding CLBP (mean 12 years, SD 11) benefit from this programme (Chapter 2-3), but the influence of duration and intensity of the programme as moderating factors remains elusive and needs to be further explored.

b) **Motivation of participants to change behaviour.** Although motivation is a selection criterion for the programme, we neither assessed this factor in a clearly valid and reproducible way at pre-treatment nor assessed it systematically over time. As ‘motivation’ is viewed as a state that is amenable to change rather than a trait that is constant [22], it is possible that those patients who were unsuccessful after having followed this programme were actually not ready or motivated to change their pain-related behaviour. Motivation or readiness to change pain-related behaviour is important for treatment compliance; it plays an important role in accomplishing and achieving treatment goals and it positively influences quality of life and may predict healthcare costs [23,24], completion of a treatment programme [25], and treatment outcome [23,26,27]. The pain Readiness to Change (RtC) model is used for conceptualising the process of adopting a self-management approach to chronic pain [22]. According to the RtC model, individuals vary in their degree of readiness to adopt a self-management approach. In Chapter 2 it is shown that behavioural change is possible and that participants adopt self-management strategies to cope with back pain complaints. The Multidimensional Pain Readiness to Change Questionnaire (MPRCQ) [28,29] is designed to measure RtC related to specific components of self-management targeted in multidisciplinary treatment programmes [30], completion of treatment, and to predict treatment outcomes [29]. However, a recent study shows that, in patients with CLBP at pre-treatment, coping behaviours instead of readiness to engage those behaviours are associated with pain-related functioning [31]. Further research is needed to assess this indicator in a clearly valid and reproducible way and to evaluate its contribution to the outcome over time.
c) **Improvement of dysfunctional cognitive behavioural factors.** It has been suggested that improvement of such factors as catastrophising cognitions (i.e. exaggeration of the threat value of pain sensations) and fear of movement behaviour might contribute to the development of CLBP [32], to persistence of CLBP [33,34], and jeopardize successful treatment outcomes [12,35-38]. This phenomenon is described in the fear avoidance model (FAM) [32]. The model postulates a causal relationship between pain catastrophising (a sign of serious injury or pathology [39]), fear of movement, disability, and experienced pain severity [32,39]. Some studies have concluded that the impact of these dysfunctional cognitive behavioural factors on outcome is diminished [40,41] or is even absent [42], which is consistent with the results presented in Chapter 4. These findings corroborate the suggestion that the sequence and relationships between pain catastrophising, fear of movement, and disability postulated in the FAM maybe different for the development of CLBP than for the recovery of disability as a result of an intervention [43]. The CPP programme has a beneficial impact on cognitive behavioural variables [44], but a closer exploration of these cognitive behavioural factors and their impact on functional ability is needed.

**Methodological considerations**

The studies presented in Chapter 2 and 3 followed an observational study design. To enhance the internal validity of the studies, several precautions in the study methodology were taken to minimise the potential influence of confounding on the outcomes. In evidence-based medicine the randomised controlled trial (RCT) is regarded as the gold standard to secure the internal validity of the study, because only the RCT is thought to resemble a true, pure experiment from which causal inference can be concluded. However, in the last decade the value of the RCT has been questioned for decisions about the use of interventions due to drawbacks of the design and available evidence from RCTs [45]. In certain situations, as in an RCT, ‘experimentation’ may be unnecessary, inappropriate, impossible, or inadequate [46]. Some of these are relevant in both surgical and conservative treatments. For example, an RCT may be inappropriate because of the random allocation and blinding procedures used. When the clinician and patient agreed upon the treatment, the clinician or patient (or both) have their preferences. This arises when the effectiveness of the intervention depends on patients’ active participation in conservative programmes, which, in turn, depends on patients’ beliefs and preferences. In (spine) surgery the results and effects also depend on surgical skills as well as surgeons’ beliefs and preferences. As a consequence, the lack of any subsequent difference in outcome between comparison groups may underestimate the benefits of the intervention. To overcome these problems, alternatives have been suggested for which the key is to use all available study designs, depending on the research question, and to perform every study with scientific rigour [47]. One of the alternatives to the RCT is routine outcome monitoring with an outcome registry, set up with an observational study design (Chapter 5). Recent spine-related studies comparing RCTs with observational study designs show comparable results [48-52], which suggests that observational study designs are complementary to RCTs.

The external validity of the study results presented in the Chapters 2-4 might be limited because we studied the prospective cohort with carefully selected patients in a secondary care setting. The conclusions of the studies performed in Chapter 2 to 4 are based on data gathered through routine outcome monitoring rather than an RCT. The outcome monitoring is performed through the web-based outcome registry of the CPP programme. The outcomes registry follows the guidelines of observational studies [53] and the recommendations listed in
Chapter 5. To further substantiate the effects found in Chapter 2, an historical controlled trial was performed to compare the magnitude of treatment effects to those of other published studies for similar populations. For this comparison an SMR was used, which is a rate ratio to compare estimates of relative treatment improvement. We were limited to the external references used to calculate these SMRs, but we found comparable effects on functional outcome between the CPP programme and surgical interventions, and beneficial effects of the CPP programme compared to less intensive treatment programmes (Chapter 2). The main characteristics of the study populations were comparable, but unmeasured discrepancies are still likely.

**Implications for clinical practice and future research**

Because of the issues discussed above regarding generalisability, implementation of this CPP programme throughout the Netherlands could be challenging and the results of implementation will need to be confirmed. Therefore, with implementation in other healthcare settings, on-site continuous outcome monitoring is needed to demonstrate continuous quality of care and to benchmark results. Cost-effectiveness studies are needed to examine the impact and the value (i.e. outcomes relative to the costs [54]) of the delivered care.

For a subgroup of participants in the CPP programme, i.e. those with previous lumbar spine surgery, the efficacy and impact of this programme compared to spine surgery is not yet clear. In the studies presented in Chapter 2-4, almost a third of the participants had previous back surgery (i.e. failed back surgery syndrome [FBSS]). To study the cost-effectiveness in this selected subgroup, a pragmatic RCT design is proposed with outcome monitoring over time. Before random assignment, two recommendations are offered: a waitlist condition to wash out previous treatment effects, and standardised management of expectations, because expectation of the outcome is predictive of a successful surgical outcome (Chapter 10).

In Chapter 4, other moderating process factors were suggested which potentially predict a successful outcome, such as a clear treatment rationale, a highly structured programme, providing a pressure-cooker model programme, the intensity or dose of treatment, and a skilful staff (e.g. the impact of the spine surgeon in the educational part of the programme). These aspects should be further explored as potential working mechanisms of the programme. Regarding the inclusion criteria, motivation or readiness to change behaviour and outcome expectations might be relevant characteristics that need to be further studied.
Theme B: Outcomes assessment

Research questions
1. What is the current value and methodology of spine outcome registries in clinical practice?
2. Which patient-related outcome measures should be used for outcomes assessment for degenerative lumbar spine disorders?
3. Which criterion can be used to define a successful outcome of interventions for patients with degenerative lumbar spine disorders?

Summary of main findings
In Chapter 5 a systematic review is presented, in which 25 spine registries from around the world were identified. No conclusions can be drawn on the value or the impact of these registries on the quality of spine care, regardless of whether the intervention was non-surgical and/or surgical. The 25 registries were heterogeneous. To improve the quality of evidence published with registry data, we presented 14 recommendations. Two recognised shortcomings are that different outcome measures are used and different core or standard outcome sets for CLBP exist. However, going forward, consensus is needed in order to standardise outcomes to be able to compare and benchmark outcomes for degenerative lumbar spine disorders. In Chapter 6 a standard set of outcomes and influencing (risk) factors for use in clinical practice and research is described. The set was compiled through a literature review and a worldwide formal (modified Delphi) consensus procedure. One of the standard outcome instruments agreed upon to assess functional ability is the Oswestry Disability Index version 2.1a (ODI) as a patient-reported outcome measure (PROM). Because in the Netherlands several informal translations exist, we translated this version of the ODI into the Dutch language according to established guidelines and then evaluated the main methodological quality properties. In Chapter 7 we describe this process and report that the Dutch version of the ODI proved to be a valid and useful tool, with good measurement properties for the assessment of functional ability amongst patients with CLBP. To indicate treatment success after spine surgery, a Patient Acceptable Symptom State (PASS), equivalent to an absolute score on the ODI version 2.1a, was estimated in Chapter 8. Using follow-up data from patients with degenerative lumbar spine registered in the Eurospine Spine Tango Spine Surgery Registry, we estimated the PASS to be 22, irrespective of the time of follow-up. We recommend using this PASS as a threshold (ODI ≤ 22) to define treatment success alongside the commonly used change-score values.

Discussion
In order to improve consistency between registries, several recommendations were made (Chapter 5). First, outcome registries should be methodologically well constructed, which requires the use of observational study methods and the identification of best practices in existing registries so that a standardised approach to registering and analysis can be achieved. This effort will depend on international collaboration and benchmarking, and will contribute to future value-based spine care.

Second, to improve the quality of spine care (i.e. outcomes of interventions; see General Introduction, 3. Outcomes of interventions for CLBP), consistent improvement strategies are needed (Chapter 5). For example, providing frequent continuous feedback (audit cycles) of outcomes captured in registries raises awareness and is recommended to enhance improvement of quality of care [55-59]. However, only two registry representatives reported...
doing monthly (or more frequent) audits (i.e. Texas Back Institute and the Dutch Spine Surgery Registry; Chapter 5). However, these registries are new and started only recently, so no results are currently available and therefore no firm conclusions can be drawn. In the past, data to compare the performance of different healthcare providers were scarce. With Sweden as a pioneer and the United States following with the National Surgical Quality Improvement Program (NSQIP) [60], which was set up to continuously monitor and enhance the quality of surgical care, clinical registries (audits) have been initiated at both regional and national levels. This has led to a demonstrable improvement in clinical outcomes and smaller variation between providers [9,58,61]. However, so far, we could not show this impact in spine care (Chapter 5). The fact that registries can have an important effect on outcomes of interventions (i.e. quality of healthcare) was reported in a study of 13 outcome registries in five countries in other medical fields (e.g. hip arthroplasty, acute myocardial infarction and cataract surgery). That study demonstrated that registries have great potential to improve health outcomes and lower healthcare costs [58]. Studies describing the relationship between improvement of outcomes and reduction of hospital costs by quality improvement programmes are scarce [62], but a recent Dutch nationwide study presented evidence for simultaneous quality improvement and cost reduction in colorectal surgery. The authors concluded that participation in a nationwide quality improvement initiative with continuous quality measurement and benchmarked feedback reveals opportunities for targeted improvements [63].

A third recommendation presented in Chapter 5 is to use a standard set of patient-related outcome measures with good measurement properties applied in a systematic approach, such as presented in Chapter 6, to make future comparisons and benchmarking possible. Currently, broad international comparisons are limited because there exists large variation in the definitions of specific diseases, in included patient-related outcomes (PROMs and clinical outcomes), and in associated influencing (risk) factors. In the literature, several recommendations have previously been published for standardised outcome measurement in low back pain research [64-67], but not specifically for use in the full cycle of care in everyday clinical practice. An outcome set for use in everyday clinical practice and for continuous improvement of the quality of spine care requires availability and validity in many languages, capacity for case-mix adjustment to ensure that comparisons are made fairly, and should focus on the outcomes that matter most to patients (Chapter 6). The proposed outcome set in Chapter 6, which was achieved with formal international consensus, seems to fulfil these requirements. Further research is needed to validate this set, including linguistic and cross-cultural translation and adaptation. In order to improve outcomes of treatment and with it the quality of spine care, research should focus on alignment of existing standard or core outcome sets with the standard set presented in Chapter 6, which is also relevant for patients and patient evaluation, and with the International Classification of Functioning, Disability, and Health (ICF) [68].

Methodological considerations

When using outcome measures, clear criteria are required for the definition of treatment success. This is a methodological challenge, as the interpretation of pre- and post-treatment scores of PROMs has been a topic of research for more than two decades [69-71] and a methodological concern remains on how best to estimate and interpret outcomes and effects of interventions. Two different concepts to define treatment success are currently commonly used and methodologically discussed: a) relative change values (i.e. relevant
change; improvement or deterioration) and b) achievement of an acceptable symptom state by reaching absolute values.

a) **Change measures** are widely recommended and commonly used in spine research. These relative measures are often referred to as a percentage baseline difference or a minimal important change (MIC) value (i.e. minimal clinical important difference [MCID] or change [MCIC]). A MIC value depicts a change, which is considered to be minimally important by patients, clinicians, or relevant others [72]. Several drawbacks concerning the use of change measures are acknowledged. For example, it is difficult to measure what is clinically relevant change to patients [73] because the definition of what is clinically important or relevant or meaningful to patients is subjective. So far, there is no agreement on what constitutes operational definitions of ‘important’ and ‘meaningful’ as applied to clinical changes [74]. Another drawback is that the achievement of change scores is highly dependent on baseline values [75-77]. Furthermore, to determine the MIC value, global perceived effect (GPE) or recovery (GPR) scales are used as the anchor (i.e. external reference). The GPE scale asks the patient to rate, on a numerical Likert scale, how much their condition has improved or deteriorated since some predefined time point. Some validity concerns are acknowledged and it might be unsuitable for use as an external reference because patients can have difficulty recalling their previous status, and their estimates of transition are biased by their current health status [78]. Finally, a change value does not indicate whether the patient is satisfied with the current state. Therefore, research into the concept of reaching absolute values is needed. An example of such an absolute measure is the achievement of an absolute score equivalent to a PASS.

b) The PASS as an **absolute value or threshold** might be a more stringent and adequate measure to indicate treatment success. A PASS might be more important to patients than change values, as it probably reflects the ultimate goal of a treatment from the patient’s perspective: ‘It’s good to feel better but it’s better to feel good.’ [79,80]. It is a measure related to the current state of the patient and it overcomes the baseline dependency as encountered for change measures. The rationale for us to use a strict absolute threshold was that patients with CLBP presenting in secondary or tertiary spine care are severely disabled. When patients with such a high ODI-value at pre-treatment assessment improve and reach a recommended change value of 10 points [81] or 30% change [81,82], they could be classified as ‘successful’ whilst in fact they are still (severely) disabled. In treatments where recovery from disability is a goal, a more stringent and absolute cut-off value, comparable to ODI values seen in ‘normal’ healthy populations, should be a measure of treatment success. The score associated with achieving PASS could be considered as a cut-off point for determining whether patients are ‘responders’ after spine surgery and this threshold could also be considered as a clinical treatment target. In this respect, treatment success would be defined as the value beyond which patients can consider themselves well [83,84] or good [80], an approach that is frequently used in the field of rheumatology [85]. In Chapter 8, we determined the PASS for the ODI to be 22 (out of 100) and used an anchor-based approach. As the external reference (anchor) the symptom-specific well-being (SSWB) Likert scale of the Core Outcome Measures Index (COMI) was used: ‘If you had to spend the rest of your life with the symptoms you have now, how would you feel about it?’ [86,87] (i.e. current status, instead of asking patients). With this approach, potential recall bias is avoided as patients are asked about their current status instead of how much their condition has changed since some predefined time point.
Implications for clinical practice and further research

Spine outcome registries can be used for continuous outcome monitoring. These registries might be useful in post-marketing surveillance for both innovative surgical and non-surgical treatments, after clinical introduction to the market, for monitoring the quality of care delivered to improve treatment outcomes, for benchmarking, and for research purposes (e.g. comparative effectiveness research). Moreover, comparing results of a strict multi-centre RCT with ‘real-life’ outcome data from routine practice can be possible if the RCT is performed within the framework of the registry, using the same questionnaires. This approach could contribute to a better understanding of the effect, both efficacy and effectiveness, of (new) interventions, and the external validity of the RCT can then be assessed.

A further recommendation proposed in Chapter 5, to improve the methodology of outcome registries, is based on the finding that various analytical approaches have been used to explain any possible differences in outcomes between spine practices. To prevent selection bias [88] and to explain real differences in outcomes between institutions, multivariate approaches with adjustment for covariates (i.e. corrections for differences in characteristics of patients treated in hospitals; ‘case-mix adjustments’) and correction for chance variation (reliability adjustments) are needed [89-92]. Owing to the fact that neither the aetiology (and thus the underlying mechanisms) nor the set of case-mix indicators of CLBP is known, any comparison of outcomes between institutions is a challenge. Advanced statistical techniques such as multilevel random-effects regression models are recommended, as these models apportion some of the variation between hospitals as being due to just chance [93,94], and need to be further explored for future use in multicentre studies comparing treatment outcomes in CLBP patients.

In the treatment of CLBP, an important outcome domain is functional ability. In the absence of objective measures and, more importantly, to incorporate patient experience, PROMs are used. To improve objectivity, measures of body function, such as mobility and muscle strength, have been used, although the correlation between these measures and the level of activity in daily life is very weak [95,96]. Promising preliminary results were obtained in studies involving continuous measurement of patients wearing activity sensors whilst performing activities of daily living [97-99]. Future research could quantify daily functioning by continuous activity monitoring and explore its use in evaluating treatments for CLBP as a complement to the commonly used PROMs.

To be able to define a successful outcome of an intervention from the patient’s perspective, further research is warranted to determine which method is beneficial. Using PASS to define treatment success, as equivalent to a ‘normal’ healthy symptom state seems promising for evaluating spine surgery for lumbar spine disorders. We think the ‘ODI ≤22’ threshold might be valuable for use in conservative treatments as well, but this needs to be further explored. The concept of PASS gets closest to the patient’s perspective, approaching ‘It’s good to feel better but it’s better to feel good.’ [79,80]. The PASS threshold is still a determination made by the clinician or researcher rather than what the patient considers to be satisfactory, whereas in evaluating treatment outcomes it is important to recognise the patient’s view as well. Anchor-based methods do not account for the risks and costs of treatment, and rarely define effects of intervention in terms of the difference in outcome with and without intervention. The Smallest Worthwhile Effect (SWE) is a concept that uses a benefit–harm trade-off...
method to determine the smallest effect that justifies the costs, risks and inconveniences of an intervention [100-102]. The SWE is defined as an important change whilst considering the burden of the intervention and focusses on whether it is worthwhile for the patient [100]. This novel method seems promising and needs to be further explored as it allows patients to weigh the benefits of treatment against the risks, costs, and inconveniences of treatment; and potentially provides estimates that are based on an intervention versus control comparison.

Another research area is the evolution of outcomes assessment with PROMs. A myriad of PROMs exist that measure functional ability in patients with CLBP [103]. Even when the standard or core outcome sets are aligned, multiple valid PROMs covering the different outcome domains are still needed for evaluating spine treatments. It is essential to explore whether it is valuable to combine PROMs into one outcome measure. A promising new approach in this area is the Patient Reported Outcomes Measurement Information System (PROMIS NIH). PROMIS is a system of reliable and precise measures of patient-reported health status for physical, mental and social well-being [104]. PROMIS scales, once validated for a specific patient group, may be calibrated and built into computer adaptive tests (CATs). A CAT integrates the advantage of measurement theory and the power of computer technology to administer a PROM that selects questions on the basis of a patient's response to previously administered questions or other prior information [104]. Highly informative questions are selected and scores are estimated that represent a person's level in a domain (e.g. physical functioning, quality of life, and pain) with the minimal number of questions without a loss in measurement precision [104,105]. It will be interesting to explore whether these tailored and individualised CATs are feasible and valid for use in future studies evaluating treatment outcomes of CLBP.
Theme C: Prediction of outcomes

Research question
1. Is it possible to develop a triage tool for CLBP, which enables valid and reliable identification of patient profiles that supports triage of the patients to a spine surgeon or to non-surgical specialists?

Summary of main findings
Chapter 9 describes the development of a decision tool for secondary or tertiary spine care specialists to decide which patients with CLBP should be seen by a spine surgeon or by non-surgical medical specialists. Based on a literature review, consensus was reached to include 47 indicators. A first version of the decision tool was developed, consisting of a web-based screening questionnaire, systematic outcomes assessment, and a provisional decision algorithm. In Chapter 10, ten patient-reported factors were found to be predictive of referral to spinal surgery amongst CLBP patients (i.e. female gender, previous back surgery, high leg pain intensity, somatisation, positive treatment expectations, being obese, having comorbidities, pain in the thoracic spine, reduced walking distance, and consultation location). The explained variance of the model was low (6%), which means that the model could only partly predict spinal surgery referral amongst CLBP patients. A longitudinal study is presented in Chapter 11, in which different patient profiles are identified for patient triage, based on pre-treatment patient-reported characteristics and one-year follow-up outcomes. Different pre-consultation profiles were determined that could predict 'response' and 'non-response' to spine surgery and to a CPP programme at follow-up. In general, the performances of all models were acceptable; in particular, the 'non-response' model to elective lumbar spine surgery performed remarkably well ($R^2$ 39%; c-index 0.83).

Discussion
Currently, a valid pre-treatment classification is lacking that accurately predicts a consistent beneficial outcome after lumbar spine surgery and non-surgical treatments [106,107]. To our knowledge, we present the first internally valid classification: a ‘proof-of-principle’ version of the Nijmegen Decision Tool for Chronic Low Back Pain (NDT-CLBP), developed for secondary care (Chapter 9-11). The current proof-of-principle tool consists of a comprehensive screening questionnaire, systematic outcome monitoring (Chapter 9), and prognostic patient-reported profiles related to treatment outcome (Chapter 11). The ultimate purpose of our research programme is to reliably identify, in two phases, patients who would most likely benefit from certain interventions. The first phase is aimed at identifying prognostic patient-reported profiles to enhance timely patient triage to a spine surgeon or a non-surgical specialist for consultation, as studied in Chapter 11. In the second phase, based on further diagnostics (e.g. imaging), the profiles will be refined to include both the indicators from the first phase and those from the diagnostic phase, to reliably refer the patient to the most appropriate treatment. This second phase is planned and part of future research, but is beyond the scope of this thesis.

International guidelines [14,15] recommend using the clinical flag approach [108] for clinical decision-making in CLBP. A diagnostic triage based on ‘red flag’ signs is recommended [14,15,109] as they are thought to be associated with a high risk of serious underlying disorders, such as infection, inflammatory disease, cancer or fracture [108,110] or nerve root disease [111].
The presence of a red flag alerts clinicians to the need for further examination and specific management [110,112,113]. Thus, we chose the red flag approach as a starting point for the preliminary version of the decision algorithm, based on multidisciplinary consensus (Chapter 9). In Chapter 11 we report that most of the patients with CLBP (92%) show at least one positive red flag but do not have a serious underlying condition. Taking the guideline recommendations literally could cause harm (e.g. unnecessary diagnostics, unnecessary exposure to radiation, unnecessary treatments, including surgery [114]), and these recommendations should be reconsidered. To evaluate indicators predicting treatment outcome, in Chapter 11 we determined different pre-treatment patient-reported profiles. The combination of indicators determined in a profile is used to predict the individual probability of treatment outcome. Owing to the high number of potential predictive indicators and the limited number of events (events per variable [EPV]), we were not able to include interaction terms (e.g. combination of red flags with yellow flags [e.g. somatisation or distress]). More detailed analyses including these interaction terms are planned when more new patients are enrolled. Moreover, we expect that combinations of red flags and clinical features, determined in the diagnostic phase, might be more informative to assist in clinical decision-making [112,113,115]. This needs to be further explored in the second phase of the NDT-CLBP when patient profiles for decision to treatment, rather than triage alone, are to be analysed and built.

**Methodological considerations**

The patient-reported profiles determined in Chapter 10 and 11 were studied and reported in line with the recommendations of the ‘prognosis research strategy’ (PROGRESS) [116] and the ‘transparent of a multivariable prediction model for individual prognosis or diagnosis’ (TRIPOD) [117]. The patient-reported profiles are based on prognostic models. The establishment of such profiles in clinical practice requires three distinct phases [118]: 1) Development (i.e. identification of important predictive indicators from an observational study); 2) Validation (i.e. testing of the profile’s predictive performance in new patients to determine whether it remains reliable and stable); and 3) Impact analysis on daily practice (i.e. assessment of the usefulness of the profile in the clinical setting to identify whether the validated profile is likely to have meaningful, beneficial consequences). Such benefits may include more accurate clinical decision-making in terms of selection and prioritisation of patients requiring intervention, improved patient outcomes, and reduced costs of care [107,119-121]. Adams et al. added an important fourth phase: Implementation (i.e. widespread acceptance and adoption of the profile in clinical practice) [120].

In the studies of Chapter 10 and 11, the models were developed and internally validated to correct for over-fit and optimism of performance measures [119,121-123]. In Chapter 10 we identified ten prognostic patient-reported indicators for referral to lumbar spine surgery. These indicators are not externally validated yet, but they are known as common predictive indicators for surgical outcome, as previous studies have identified them, too (e.g. [124-126]). This (preliminary) result gave us the confidence and arguments to further explore and examine the patient-reported profiles longitudinally, based on treatment outcomes (Chapter 11). Due to the observational study design used in Chapter 11, confounding by indication might be present. Confounding by indication is where allocation to treatment is subject to a black box of an unrecognised or unmeasured process associated with those who are treated, which is guided by the experience and preferences of clinicians who use their expert judgment to decide whether to treat a patient [127]. Concealed randomisation has been suggested to interfere in
Summary and General discussion

the relationship of prognosis and prescription [128]. Due to patient and clinician preferences, and as lumbar spine surgery and CPP programmes are regarded as mutually exclusive, randomisation is not feasible. We therefore described the source population and both the cohorts and performed subgroup analyses to discover patterns and the patient profiles. As mentioned above, owing to the relatively high amount of potential predictive variables and relatively few events, expressed in EPV, we were not able to introduce interaction terms. As such, we cannot rule out that patients might actually fit into both profiles, e.g. patients who responded well to spine surgery might have responded well in the CPP programme as well. However, one should bear in mind that the patient profiles are used for decision-making in patient triage to a surgical or non-surgical specialist instead of the actual treatment. Future studies are needed to validate the prognostic profiles using interaction terms for type of treatment, and to further substantiate the decision process.

Implications for clinical practice and further research

Before widespread implementation and use of the patient profiles for triage to a surgical or non-surgical specialist, as determined in Chapter 11, the underlying prognostic models need to be further validated in new patients and in other, similar secondary spine practices, and impact studies exploring the cost-effectiveness and the clinical usefulness are needed. Including more patients with follow-up outcomes would allow studying interaction effects between combinations of potential prognostic indicators and interaction effects between types of treatment, which could be used to further refine the prognostic profiles. In the study of Chapter 11, we analysed and described patient-reported profiles for elective lumbar spine surgery and multidisciplinary bio-psychosocial treatment (i.e. CPP programme). Further research is planned to explore patient profiles for other treatments, e.g. invasive pain management or ‘no treatment’ (i.e. counselling and physical therapy in primary care), to explore whether patients with features of axial spondylarthropathy could be determined by their profile for early referral to rheumatologists for further diagnostics and treatment, and to conduct a pilot study to determine whether the triage tool is feasible for use in primary care.

To develop the patient profiles (Chapter 11), we examined associations between potential prognostic indicators and one outcome domain: functional ability. Strict evidence-based absolute thresholds were used to define the outcome. As yet, no consensus exists on how to define the outcome and which measure, relative or absolute change, to use to define success or failure. To be able to define ‘response’ and ‘non-response’ after treatment for CLBP, clear evidence-based and expert consensus-based criteria, including the measures that operationalise these definitions, are needed for ‘responder analysis’ to make future comparisons in both research and clinical practice possible. To define these criteria a research approach comparable with the Osteoarthritis Research Society International (OARSI) set of responder criteria for osteoarthritis [129] could be used.

The triage profiles of the NDT-CLBP belong to the first phase of the decision-making process for a certain treatment. The next, second, diagnostic phase contributes to the final decision between clinician and patient as to which treatment is most appropriate. Future research should examine whether (combinations of) biomedical and psychosocial indicators that were not included in Chapter 11 should be incorporated into the models, to extend and further refine the different patient profiles. For this, relevant data of the diagnostic phase (e.g. specific diagnosis [e.g. spondylolisthesis spinal stenosis], physical characteristics) should be added to
Summary and General discussion

the currently developed (patient-reported) triage tool. Growing interest exists to determine clinical phenotypes. Phenotypes refer to observable traits of an individual organism [130], consisting of biomedical and psychosocial factors. Based on novel techniques, preliminary evidence suggests that radiographic characteristics (Modic changes), biomarkers, and genetic profiles, might possibly explain CLBP bio-medically [131]. In line with the recommendation of Samartzis et al. [131], a global consortium (i.e. AO Personalized Spine Care consortium) has been set up, consisting of a multidisciplinary group of spine and pain specialists and researchers, to further refine the understanding of CLBP and to further study both the decision tool for patient triage and for treatment to enhance stratified and personalised spine care.

Towards a paradigm shift in CLBP, away from stepped care in secondary spine care?

Studying different prognostic patient profiles based on the NDT-CLBP might contribute to further refine and support future treatment decisions in CLBP. CLBP is multifactorial in nature and the underlying mechanisms remain largely unknown. As yet aetiological studies have contributed little to diagnostics in CLBP and corresponding treatments that are ultimately expected to lead to successful treatment outcomes. Although future studies may reveal some aetiological factors, we speculate that a shift from classic causal reasoning towards reasoning in prognostic patient profiles may form an alternative paradigm. Patient profiles may contribute meaningfully to the optimisation of decision-making in the treatment of CLBP, even without full knowledge of the aetiology of CLBP.

If the developed patient profiles are externally validated, the impact is shown and if they are proven to be of value in decision making (i.e. they can predict which patient will benefit from which treatment, even without a well-defined cause), then the recommendations in the (inter-) national guidelines will need to be reconsidered. In the National Institute for Health and Care Excellence (NICE) guideline [15] as well as in a recently released draft version of the Dutch national guideline for instrumented lumbar spine surgery [132], a stepped care approach is suggested: before lumbar spine surgery is considered, a CPP programme is indicated (consensus; low quality evidence). Although further research is needed to substantiate and validate the patient profiles developed in Chapter 11, the different profiles suggest that different subgroups of patients could be identified in the heterogeneous CLBP population that would benefit from treatments. This would mean — instead of the recommendation made in the guidelines for a serial (stepped care) approach, yielding spine surgery as an 'end-of-line' treatment — that a more parallel approach might be feasible. For example, some patients may be identified who will benefit from surgery, without having undergone a full non-operative cycle of care, and some patients should never undergo surgery, despite failing all non-operative treatments. Then, a remarkable paradigm shift in clinical reasoning of medical specialists in spine care would be required.
Concluding remarks

The CLBP population is heterogeneous and the condition lacks diagnostic clarity. The failure to differentiate between underlying causes is one of the reasons why a plethora of invasive and non-invasive interventions exist for the same symptom. Despite decades of research and improved quality of clinical trials, the reality is that the treatments offered to patients have led to inconsistent results. Based on the research in this thesis, it seems that:

1. The introduced CPP programme is effective for a selected subgroup of patients with CLBP.
2. Using strict selection criteria for patient entrance to interventions, clear outcome definitions and continuous outcome monitoring of treatments using an outcome registry can contribute to improving the quality of delivered spine care.
3. For the definition of treatment success in secondary spine care, the absolute ODI-22 threshold is recommended for use alongside the commonly used change measures.
4. Different subgroups within the heterogeneous CLBP population are evident, based on different prognostic patient-reported profiles and response or non-response after treatments in secondary care. To our knowledge the developed prognostic patient-reported profiles of the NDT-CLBP are the first that seem to recognise these more homogeneous subgroups.
5. The non-response prognostic patient-reported profile for elective lumbar spine surgery performed remarkably well. This suggests that patients with a high probability for persisting severe disability after surgical intervention could be identified even before actual consultation, so that unnecessary and unhelpful surgery could be avoided.
6. In the future, when these patient profiles have been shown to remain stable after validation in new patients and in other secondary spine practices, they may be used to enhance timely patient triage to the right surgical or non-surgical specialist and contribute to the decision-making between clinician and patient.

Subgrouping CLBP based on treatment outcomes still needs to be a research priority, and further high-quality research is needed in the screening and diagnostics of CLBP. Exploring the validity and the impact of the patient profiles presented in this thesis is required before firm conclusions can be drawn regarding the NDT-CLBP. Including the diagnostics of the second phase of the NDT-CLBP in the profiles of the decision tool might contribute to improvement of treatment outcomes. These profiles could lead to a paradigm shift in clinical reasoning and decision-making and ultimately to a more efficient use of healthcare resources and reduction of the tremendous worldwide burden of low back pain.
References


11. Wand BM, O’Connell NE. Chronic non-specific low back pain - sub-groups or a single mechanism? BMC. Musculoskelet.Disord. 2008;9:11


60. NSQIP. The National Surgery Quality Improvement Program (NSQIP) [Internet]. American College of Surgeons, Chicago 2016 [cited 2016 July 20]; Available from: https://www.facs.org/quality-programs/acsnsqip


70. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual.Life Outcomes. 2006;4:54


72. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual.Life Outcomes. 2006;4:54


77. de Vet HC, Foumani M, Scholten MA, Jacobs WC, Stiggelbout AM, Knol DL et al. Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. J.Clin.Epidemiol. 2015;68:518-524
79. Dougados M. It's good to feel better but it's better to feel good. J.Rheumatol. 1976;33:90-94


120. Adams ST, Leveson SH. Clinical prediction rules. BMJ 2012;344:d8312
132. CBO, Nederlandse Orthopaedische Vereniging. Concept Richtlijn Geinstrumenteerde spinaalchirurgie bij degeneratieve aandoeningen van de thoracolumbale wervelkolom [Guideline in Dutch]; 2015