What can we learn from long-term studies on chronic low back pain?
A scoping review

Alisa L. Dutmer1 · Remko Soer1,2,3 · André P. Wolff3,4 · Michiel F. Reneman1 · Maarten H. Coppes5 · Henrica R. Schiphorst Preuper1,3

Received: 9 June 2021 / Revised: 22 December 2021 / Accepted: 5 January 2022 / Published online: 19 January 2022 © The Author(s) 2022

Abstract

Purpose A scoping review was conducted with the objective to identify and map the available evidence from long-term studies on chronic non-specific low back pain (LBP), to examine how these studies are conducted, and to address potential knowledge gaps.

Method We searched MEDLINE and EMBASE up to March 2021, not restricted by date or language. Experimental and observational study types were included. Inclusion criteria were: participants between 18 and 65 years old with non-specific sub-acute or chronic LBP, minimum average follow-up of > 2 years, and studies had to report at least one of the following outcome measures: disability, quality of life, work participation, or health care utilization. Methodological quality was assessed using the Effective Public Health Practice Project quality assessment. Data were extracted, tabulated, and reported thematically.

Results Ninety studies met the inclusion criteria. Studies examined invasive treatments (72%), conservative (21%), or a comparison of both (7%). No natural cohorts were included. Methodological quality was weak (16% of studies), moderate (63%), or strong (21%) and generally improved after 2010. Disability (92%) and pain (86%) outcomes were most commonly reported, followed by work (25%), quality of life (15%), and health care utilization (4%). Most studies reported significant improvement at long-term follow-up (median 51 months, range 26 months–18 years). Only 10 (11%) studies took more than one measurement > 2 year after baseline.

Conclusion Patients with persistent non-specific LBP seem to experience improvement in pain, disability and quality of life years after seeking treatment. However, it remains unclear what factors might have influenced these improvements, and whether they are treatment-related. Studies varied greatly in design, patient population, and methods of data collection. There is still little insight into the long-term natural course of LBP. Additionally, few studies perform repeated measurements during long-term follow-up or report on patient-centered outcomes other than pain or disability.

Keywords Scoping review · Chronic low back pain · Long-term follow-up · Disability · Quality of life · Work participation
Introduction

Low back pain (LBP) is very common and poses a great health risk for society. Worldwide, it is the number one cause of years lived with disability [1]. Up to 84% of the population will experience LBP at least once during their lifetime [2]. In roughly 90% of cases, a specific source for the LBP cannot be identified [3]. LBP is strongly associated with disability [1, 4], work absence [5, 6], and reduced quality of life [6, 7]. As a result, medical and particularly non-medical costs related to LBP are very high [8, 9].

Most patients improve substantially in the first six weeks after the onset of LBP [10]. However, one year after onset, approximately two thirds of patients still experience pain and disability [10–12]. Currently, LBP is looked at more and more as a long-lasting or recurrent condition rather than a series of unrelated episodes [9, 13]. A review on the long-term course (follow-up ranged from one to 28 years) of LBP in the general population found that most patients experienced a somewhat stable or fluctuating occurrence of LBP over time [14]. Becoming pain free was never reported as a common finding.

Despite the effects of LBP on physical, psychological, and social well-being, there are few longitudinal studies reporting multiple patient-centered outcomes. Cohort studies with long-term follow-up (> 2 years) often confine to investigating the presence of pain (yes/no) or the number of days with pain over the past month(s) or year [13, 14]. Several consensus statements have been published on outcome measures in chronic (back) pain research [15–17]. Most reports specifically provide recommendations for the evaluation of clinical trials, but there is an overall understanding that reporting on pain alone in LBP research is insufficient. Other important outcome domains include measures of physical function, generic measures of health and well-being, quality of life, and work (dis)ability.

At present, it is unclear what evidence is available from long-term studies on chronic non-specific LBP. More specifically, from studies examining patient-centered outcomes other than pain. We conducted a scoping review with the objective to identify and map the available evidence from studies on chronic LBP with long-term follow-up, to examine how these studies are conducted, and to address potential knowledge gaps. Where systematic reviews typically focus on more narrow and well-defined questions with appropriate study designs chosen in advance, a scoping review tends to address broader topics where many different study designs might be applicable [18]. For the present study, we included experimental and observational studies reporting at least two-year follow-up on disability, quality of life, work participation or health care utilization in patients with chronic non-specific LBP. The results are not intended to provide evidence to inform clinical practice, but rather to gain insight into the scientific literature that is currently available. For studying the feasibility, appropriateness or effectiveness of a certain treatment or practice, a systematic review is a more valid approach [19].

Methods

The PRISMA Extension for scoping reviews (PRISMA-ScR) was used as a reporting guideline for this review [20]. Although critical appraisal is optional, for the present study we evaluated methodological quality of the included studies with a quality assessment tool in order to be able to address any potential gaps in the literature related to low quality of research [21].

Eligibility criteria

Types of studies

Both experimental and observational studies investigating non-specific LBP with baseline measures and a minimum (mean) follow-up of > 2 years were included. Case reports and review studies were excluded.

Participants

Study participants were adults with sub-acute (6–12 weeks) or chronic (> 12 weeks) non-specific low back pain at study baseline, with or without leg pain. The average age of the study population had to be between 18 and 65 years. Studies that reported on LBP due to a specified physical cause (e.g., infection, tumor, osteoporosis, fracture, structural deformity, inflammatory disorder, radicular syndrome or cauda equina syndrome) were excluded. Studies on patients with LBP due to failed back surgery syndrome (FBSS) and LBP due to degenerative changes such as disk degeneration, osteoarthritis of facet joints, and a grade 1 degenerative spondylolisthesis were included, provided that there were no neurological symptoms. Little to no association has been found between imaging findings of these types of spine degeneration and the presence of LBP [22–26]. We therefore classified these (radiological) diagnoses as non-specific. Studies with mixed LBP groups (specific and non-specific cause for LBP) or mixed pain populations (e.g., neck pain and LBP) were excluded unless subgroup data for baseline and follow-up were presented.
Outcome measures

To be included, studies had to report on at least one of the following outcome measures: disability, quality of life, work participation, or health care utilization. Pain was also an outcome measure, but studies that only reported on pain were not included.

Search methods for identification of studies

Electronic searches in MEDLINE and EMBASE were conducted using indexed terms and free text words. The searches were not restricted by date, language, or place of publication. The search strategy included terms related to LBP, long-term follow-up and outcome measures (Supplementary Digital Content [SDC] 1). The search results for both databases were downloaded into RefWorks and duplicates were removed. An initial literature search was performed, followed by several updates, of which the last took place on March 5 2021. Initially, search terms for spondylolysis and spondylolisthesis were included. However, these were removed in the updated searches. We found that studies that were retrieved with these search terms (and that would not have been retrieved by searching for terms related to low back pain) targeted patients with spondylolisthesis with a higher than grade 1 degree of severity.

Data collection and analysis

Study selection

Three review authors independently screened titles, abstracts, and full text of the studies retrieved from the databases. One author (AD) screened all studies and two authors (RS, RSP) each screened half. The inclusion criteria included type of participants, length of follow up, and outcome measures. To determine interrater agreement, a sample of 200 studies was selected for the three reviewers to screen on title, abstract and full text. Agreement ranged from 98 to 99% between reviewers with kappa scores ranging 0.56–0.98 (moderate or substantial agreement). However, kappa scores are deemed not very reliable for ‘rare findings’ [27] and in this sample of 200 studies ultimately only 3 studies were included after consensus was reached. Any disagreement in the selection of studies was discussed until consensus was reached. If the three reviewers could not reach consensus, the fourth reviewer (MR) was consulted.

Quality assessment

The Effective Public Health Practice Project Quality Assessment Tool (EPHPP) was used to evaluate methodological quality of the studies [28]. The tool can be used to evaluate a variety of study designs such as RCTs, observational, cross sectional, and before-and-after studies. The EPHPP assesses six domains: (1) selection bias, (2) study design, (3) confounders, (4) blinding, (5) data collection method, and (6) withdrawals and dropouts. Each domain can be rated strong, moderate, or weak resulting in a global rating of strong (no weak ratings), moderate (one weak rating), or weak (two or more weak ratings) for each study. The confounders domain was scored ‘not applicable’ when there was no comparison or control group, since the corresponding question was phrased “Were there important differences between groups prior to the intervention?”. Content and construct validity of the EPHPP have been established and inter-rater reliability is fair for the individual domains (ICC = 0.60) and excellent for the global rating (ICC = 0.77) [28, 29]. Four reviewers assessed methodological quality of the studies. One author (AD) assessed all studies and three authors (RS, RSP, MR) each assessed one third of the studies. Disagreements were resolved between the authors assessing the study or when in doubt were discussed with all four assessing authors to reach consensus.

Data extraction and synthesis

The following data were extracted by one author (AD) from each paper and presented in supplementary tables (SDC 2): first author, study setting and country, study design, intervention(s), patient characteristics (diagnoses, age, % female), outcome domain(s), instrument(s), duration of follow-up, and results of measurements taken at baseline and > 2 year follow-up. This includes the results of any responder analyses (i.e., the proportion of patients achieving a pre-defined level of improvement) [30]. For randomized controlled trials (RCT), results from the intention-to-treat analyses were reported. Studies were organized thematically according to intervention type. Study characteristics were also summarized in a narrative format and the overall findings were presented in a summary table. Per outcome, the number of treatment arms that showed a significant (p < 0.05) improvement, decline, or no change compared to baseline was reported. The number of studies that did not report p-values for the change in outcome at follow-up was also reported.

Results

Study selection

Together, the initial and updated searches returned 10,312 articles, of which 90 ultimately met the inclusion criteria (Fig. 1). Follow-up results of one study were presented in two different articles [31, 32]. An overview of study
characteristics can be found in SDC 2. Studies \((n = 89)\) were classified according to the type of treatment(s) that was investigated: invasive \((72\%,\ n = 64;\) Table 1, SDC 2), conservative \((21\%,\ n = 19;\) Table 2, SDC 2), or a comparison of invasive and conservative treatments \((7\%,\ n = 6;\) Table 3, SDC 2). By definition, (minimal) invasive procedures require (1) a method of access to the body (incision, natural orifice, or percutaneous access), (2) instrumentation (e.g., endoscopes, catheters, scalpels), and (3) requirement for operator skill [33]. All non-invasive treatments were classified under conservative treatments.

### Quality assessment

Global quality rating was weak for 14 \((16\%\), moderate for 56 \((63\%\), and strong for 19 \((21\%\) studies (Table 1). A global weak rating was more common with studies published before 2010, while most studies that rated strong were published in the last decade (Fig. 2). Most common design was either a prospective \((44\%,\ n = 39)\) or retrospective cohort study \((31\%,\ n = 28)\) (both rated ‘moderate’). Twenty-one studies \((24\%)\) conducted an RCT and one study was classified as a controlled clinical trial [34] (both rated ‘strong’). Weak ratings were prevalent with the domain ‘selection bias’, while strong ratings were prevalent for ‘data collection method’. Studies rated predominantly moderate \((42\%,\ n = 37)\) or strong \((44\%,\ n = 39)\) on ‘withdrawals and dropouts’. Sixty

### Study Characteristics

#### Year published

Studies were published between 1985 and 2021, with 52 out of 89 studies \((58\%)\) published in the last decade (Fig. 2).

#### Study Setting

The majority of selected studies \((83\%,\ n = 74)\) were from Western countries (SDC 2). More specifically, from European countries \((54\%,\ n = 48)\), such as Germany \((10\%,\ n = 8)\), Sweden, the UK \((both\ 9\%,\ n = 7)\), Norway, the Netherlands \((both\ 8\%,\ n = 7)\), and from the USA \((27\%,\ n = 24)\). Thirteen studies \((15\%)\) were from Asian countries of which seven \((8\%)\) from China. Two studies were from Brazil \((2\%)\). There were no studies from African countries, Central America, or Eastern Europe.

Less than half of the selected studies \((44\%,\ n = 39)\) specified the setting in which they took place. Forty-six out of 64 studies on invasive treatments did not report or were unclear in their report on where a specific intervention took place. The 18 remaining studies \((20\%)\) specified they took place in (university) hospitals or (out-patient) medical practices. Studies on conservative treatments mostly took place in (university) hospitals, physiotherapy clinics, and chiropractic and general practices. Five out of six studies that compared invasive with conservative treatments took place in university hospitals.

#### Interventions

Most common types of invasive treatment were lumbar fusion \((38\%\ of\ studies,\ n = 34)\ and disc arthroplasty \((25\%,\ n = 22)\), followed by intradiscal therapies (e.g., intradiscal electrothermal therapy or intradiscal bone marrow injection; \(11\%,\ n = 10)\, and implantable therapies (e.g., spinal cord stimulation) [35, 55, 86] (SDC 2). Less common were interspinous process devices [39, 63], dynamic spine stabilization systems [57, 85], and basivertebral nerve ablation [48]. Two studies used sham infiltration as a control for intradiscal bone marrow injection [36, 70].

Most common conservative treatments were multidisciplinary treatment \((10\%\ of\ studies,\ n = 9)\), physiotherapy or exercise training \((7\%,\ n = 6)\, cognitive therapies \((4\%,\ n = 4)\), advice and/or education \((4\%,\ n = 4)\). Other treatments consisted of (non-operative) care as usual [108, 112, 121].
Table 1  EPHPP quality assessment scores of the included studies

| References                     | EPHPP domain |
|--------------------------------|--------------|
|                                | Selection bias | Study design | Confounders | Blinding | Data collection method | Withdrawals and dropouts | Global rating |
| Studies on invasive treatments |              |              |              |          |                      |                          |              |
| Al-Kaisy et al. [35]           | M            | M            | n/a          | M        | S                     | S                         | Strong        |
| Amirdelfan et al. [36]         | M            | S            | W            | M        | S                     | M                         | Moderate       |
| Aunoble et al. [37]            | M            | M            | n/a          | M        | S                     | S                         | Strong        |
| Axelsson et al. [38]           | W            | M            | n/a          | M        | W                     | M                         | Weak          |
| Buric and Pulidori [39]        | W            | M            | n/a          | M        | S                     | S                         | Moderate       |
| Burkus et al. [40]             | M            | M            | W            | M        | S                     | W                         | Weak          |
| Buttermann and Mullin [31] and Buttermann et al. [32] | M | S | M | S | S | S | Strong |
| Buttermann et al. [41]         | M            | S            | S            | S        | S                     | S                         | Strong        |
| Cakir et al. [42]              | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Cheng et al. [43]              | W            | M            | n/a          | M        | S                     | S                         | Moderate       |
| Cheng et al. [44]              | M            | M            | n/a          | M        | S                     | M                         | Strong        |
| Chung et al. [45]              | W            | S            | S            | M        | S                     | S                         | Moderate       |
| Corenman et al. [46]           | W            | M            | n/a          | M        | S                     | S                         | Moderate       |
| Di Silvestre et al. [47]       | W            | M            | S            | M        | S                     | M                         | Moderate       |
| Fischgrund et al. [48]         | W            | M            | n/a          | M        | S                     | S                         | Moderate       |
| Formica et al. [49]            | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Geerdes et al. [50]            | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Gepstein et al. [51]           | W            | M            | n/a          | M        | S                     | S                         | Moderate       |
| Gioia et al. [52]              | W            | M            | n/a          | M        | S                     | S                         | Weak          |
| Gornet et al. [53]             | M            | S            | S            | M        | S                     | M                         | Strong        |
| Guyer et al. [54]              | M            | S            | S            | M        | S                     | W                         | Moderate       |
| Hamm-Faber et al. [55]         | W            | M            | n/a          | M        | S                     | S                         | Moderate       |
| Houten et al. [56]             | M            | M            | n/a          | M        | W                     | W                         | Weak          |
| Kareem and Ulbricht [57]       | M            | M            | n/a          | M        | S                     | W                         | Moderate       |
| Katsimihias et al. [58]        | W            | M            | n/a          | W        | S                     | M*                        | Weak          |
| Kuslich et al. [59]            | W            | M            | n/a          | M        | W                     | S                         | Moderate       |
| Lee et al. [60]                | M            | M            | n/a          | M        | S                     | S                         | Strong        |
| Liang et al. [61]              | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Lu et al. [62]                 | W            | M            | n/a          | M        | S                     | S                         | Moderate       |
| Lu et al. [63]                 | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Madan and Boeree [64]          | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Madan et al. [65]              | W            | M            | S            | M        | S                     | M                         | Moderate       |
| Maestrelli et al. [66]         | W            | M            | n/a          | M        | S                     | S                         | Moderate       |
| Malham and Parker et al. [67]  | M            | M            | n/a          | M        | S                     | S                         | Strong        |
| Meir et al. [68]               | W            | M            | n/a          | M        | S                     | S                         | Moderate       |
| Niemeyer et al. [69]           | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Noriega et al. [70]            | W            | S            | W            | S        | S                     | S                         | Weak          |
| Nunley et al. [71]             | M            | M            | n/a          | M        | S                     | W                         | Moderate       |
| Nystrom et al. [72]            | M            | M            | n/a          | M        | S                     | S                         | Strong        |
| Ohtori et al. [73]             | W            | S            | S            | M        | S                     | W                         | Weak          |
| Pan et al. [74]                | M            | M            | n/a          | M        | S                     | S                         | Strong        |
| Park et al. [75]               | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Park et al. [76]               | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Peng et al. [77]               | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Petillon et al. [78]           | W            | M            | n/a          | M        | S                     | M                         | Moderate       |
| Pettine et al. [79]            | M            | M            | n/a          | M        | S                     | S                         | Strong        |
Table 1 (continued)

| References                        | EPHPP domain                        | Selection bias | Study design | Confounders | Blinding | Data collection method | Withdrawals and dropouts | Global rating |
|-----------------------------------|-------------------------------------|----------------|--------------|-------------|----------|-------------------------|---------------------------|---------------|
| Pihlajamaki et al. [80]           |                                     | W              | M            | n/a         | M        | W                       | M                         | Weak          |
| Pimenta et al. [81]               |                                     | W              | S            | S           | M        | S                       | S                         | Moderate      |
| Pimenta et al. [82]               |                                     | W              | M            | n/a         | M        | S                       | S                         | Moderate      |
| Plais et al. [83]                 |                                     | W              | M            | n/a         | M        | S                       | M                         | Moderate      |
| Pokorny et al. [84]               |                                     | W              | M            | n/a         | M        | S                       | S                         | Moderate      |
| Putzier et al. [85]               |                                     | W              | S            | S           | M        | S                       | M                         | Moderate      |
| Raphael et al. [86]               |                                     | W              | M            | n/a         | M        | W                       | M                         | Weak          |
| Ren et al. [87]                   |                                     | W              | M            | n/a         | M        | S                       | M                         | Moderate      |
| Rouben et al. [88]                |                                     | W              | M            | n/a         | M        | S                       | M                         | Moderate      |
| Saal and Saal [89]                |                                     | W              | M            | n/a         | M        | S                       | S                         | Moderate      |
| Schimmel et al. [90]              |                                     | W              | M            | n/a         | M        | S                       | M                         | Moderate      |
| Schulte et al. [91]               |                                     | W              | M            | n/a         | M        | S                       | M                         | Moderate      |
| Sieve et al. [92]                 |                                     | W              | M            | n/a         | M        | S                       | S                         | Moderate      |
| Sköld et al. [93]                 |                                     | M              | S            | W           | M        | S                       | S                         | Moderate      |
| Strupe et al. [94]                |                                     | W              | M            | n/a         | M        | S                       | M                         | Moderate      |
| Thalgott et al. [95]              |                                     | W              | M            | n/a         | M        | S                       | M                         | Moderate      |
| Wuertinger et al. [96]            |                                     | W              | M            | n/a         | M        | S                       | W                         | Weak          |
| Zeilstra et al. [97]              |                                     | W              | M            | n/a         | M        | S                       | W                         | Weak          |

Studies on conservative treatments

| References                        | EPHPP domain                        | Selection bias | Study design | Confounders | Blinding | Data collection method | Withdrawals and dropouts | Global rating |
|-----------------------------------|-------------------------------------|----------------|--------------|-------------|----------|-------------------------|---------------------------|---------------|
| Bendix et al. [98]                |                                     | M              | S            | M           | M        | W                       | M                         | Moderate      |
| Bentsen et al. [99]               |                                     | W              | S            | W           | M        | W                       | S                         | Weak          |
| Carvalho et al. [100]             |                                     | W              | M            | n/a         | M        | S                       | S                         | Moderate      |
| Groot et al. [101]                |                                     | W              | M            | n/a         | M        | S                       | S                         | Moderate      |
| Haas et al. [102]                 |                                     | M              | M            | S           | M        | S                       | W                         | Moderate      |
| Hamre et al. [103]                |                                     | S              | M            | n/a         | M        | S                       | M                         | Strong        |
| Indahl et al. [34]                |                                     | M              | S            | S           | M        | S                       | S                         | Strong        |
| Lamb et al. [104]                 |                                     | M              | S            | S           | M        | S                       | W                         | Moderate      |
| Lanes et al. [105]                |                                     | W              | M            | n/a         | M        | W                       | M                         | Weak          |
| Lankhorst et al. [106]            |                                     | W              | M            | n/a         | M        | S                       | M                         | Moderate      |
| Patrick et al. [107]              |                                     | S              | M            | n/a         | M        | S                       | W                         | Moderate      |
| Peng et al. [108]                 |                                     | W              | M            | n/a         | M        | S                       | S                         | Moderate      |
| Raaq et al. [109]                 |                                     | W              | M            | n/a         | M        | S                       | S                         | Moderate      |
| Rantonen et al. [110]             |                                     | M              | S            | S           | M        | S                       | S                         | Strong        |
| Rasmussen-Barr et al. [111]       |                                     | M              | S            | S           | M        | S                       | M                         | Strong        |
| Rhyne et al. [112]                |                                     | W              | M            | n/a         | M        | S                       | M                         | Moderate      |
| Udby et al. [113]                 |                                     | M              | M            | S           | M        | S                       | S                         | Strong        |
| Van Hoof et al. [114]             |                                     | W              | M            | n/a         | M        | S                       | S                         | Moderate      |
| Vibe Fersum et al. [115]          |                                     | S              | S            | S           | W        | S                       | W                         | Weak          |

Studies comparing invasive and conservative treatments

| References                        | EPHPP domain                        | Selection bias | Study design | Confounders | Blinding | Data collection method | Withdrawals and dropouts | Global rating |
|-----------------------------------|-------------------------------------|----------------|--------------|-------------|----------|-------------------------|---------------------------|---------------|
| Brox et al. [116]                 |                                     | M              | S            | S           | M        | S                       | S                         | Strong        |
| Froholdt et al. et al. [117]      |                                     | M              | S            | S           | M        | S                       | S                         | Strong        |
| Froholdt et al. [118]             |                                     | S              | S            | S           | M        | S                       | M                         | Strong        |
| Furunes et al. [119]              |                                     | M              | S            | W           | M        | S                       | S                         | Moderate      |
| Hedlund et al. [120]              |                                     | M              | S            | S           | M        | S                       | S                         | Strong        |
| Kleimeyer et al. [121]            |                                     | W              | M            | S           | M        | S                       | M                         | Moderate      |

EPHPP Effective Public Health Practice Project Quality Assessment Tool, S strong, M moderate, W weak, n/a not applicable

*a Withdrawals and dropout rating varies per time of measurement: strong at 3 years follow-up (< 20% dropouts), moderate at 4 years (< 40% dropouts), and weak at 5, 6, and 7 years follow-up (> 40% dropouts)*
chiropractic care or primary care by a medical doctor [102], anthroposophic medicine [103], rehabilitation treatment [109], or open label placebo pills [100].

With the exception of two control groups that were assessed in studies on conservative treatments [98, 110], there were no studies examining long-term outcomes of LBP in people receiving no treatment. Two studies reported examining the natural history of LBP; however, their patient samples completed Swedish Back School [106] or received two months of conservative treatment [108] and were therefore categorized under ‘conservative treatments’ in this review.

Patient characteristics

Selection criteria of this review were set to include only studies on adults with sub-acute or chronic non-specific LBP. This also included patients with LBP due to FBSS, or degenerative changes such as disk degeneration and grade 1 spondylolysisis, provided that there were no neurological symptoms. One study exclusively included patients with sub-acute LBP [34] and five studies included both patients with sub-acute and CLBP [102–104, 110, 111].

The majority of studies (91%, n = 64) on invasive treatments (with or without conservative treatment as a control) included patients that fit their criteria for either degenerative disc disease (DDD), discogenic pain, internal disc disruption or a combination thereof. Other studies selected patients with Modic type 1 or 2 changes [48], patients with CLBP and radiating pain to the lower limb(s) [52], FBSS [55], either FBSS or mechanical LBP [86], or LBP originating from the endplate [77].

Only two studies investigating conservative treatment options sought to include patients with discogenic pain [108, 112]. One study specifically excluded patients with disk degeneration [100]. Commonly, patients with CLBP (58%, n = 11), sub-acute LBP [34], or both sub-acute and CLBP (29%, n = 5) were eligible for inclusion. Added criteria were: still working [111], permanent employment [110], or sick-leave due to LBP [34, 107]. One study reported results separately for patients with CLBP with or without modic changes [113].

Outcomes measurements

For the selected studies, disability (92%, n = 82) and pain (86%, n = 77) were the most commonly measured outcome domains, followed by work (25%, n = 22), and quality of life (15%, n = 13) (SDC 2). Only four studies (4%) measured health care use [85, 99, 101, 114]. Five out of seven most frequently used outcome measures were patient reported outcome measures (PROMs) of pain and disability (Fig. 3). The Oswestry Disability Index (ODI), Oswestry Disability Index; VAS, Visual Analogue Scale; SF-36, Short Form (36) Health Survey; NRS, Numerical Rating Scale; RMDQ, Roland Morris Disability Questionnaire.
Follow-up

Follow-up ranged between 26 months and 18 years with a median of 51 months (SDC 2). Forty-three studies (48%) reported an (average) duration of follow-up between 24 and 48 months, 22 (25%) studies between 49 months and six years, and 24 (27%) studies over six years. Only ten studies (11%) took more than one measurement at > 2 year after baseline. Follow-up was available for > 80% of patients in 39 studies, between 60 and 80% in 12 studies, and < 60% in six studies. The percentage was unclear in six studies. The remaining 26 studies were retrospective studies that included patients based on complete availability of follow-up. Furthermore, a total of 36 studies (all 28 retrospective studies, seven prospective studies, and one RCT [98]) reported only baseline results of those patients that completed a minimum length of follow-up.

Responder analyses

Twenty-six out of 89 studies (29%) reported the results of a responder analysis; 23 studies on invasive treatments, two studies on conservative treatments and one study that compared invasive with conservative treatments (SDC 2). An improvement in disability, measured with the ODI, was most commonly used to determine clinical success (85%, n = 22), followed by an improvement in back pain or leg pain (38%, n = 10) measured with VAS or NRS. The cut-off for clinical success varied greatly per instrument; 10 different cut-offs were used for the ODI and seven for the VAS or NRS. One study reported clinical success on pain and disability using an improvement in subscales for pain and functioning of the SF-36 [89]. Other studies analyzed improvement in quality of life (SF-36 Physical Component Scale) [67] or improvement in both pain and disability [44, 48, 60].

Summary of findings at long-term follow-up

Table 2 summarizes the overall findings of the selected studies per treatment type and duration of follow-up. Reported results were not specified for diagnoses or disease characteristics. Per outcome, the number of treatment arms that showed a significant improvement (‘+’), no significant change (‘0’), or a significant decline compared to baseline (‘−’) was reported. Several studies did not report p-values for the change in outcome at follow-up (‘?’). Results on work related outcomes were very rarely reported with a statistical level of significance. However, almost all results without a reported p-value showed some level of improvement between baseline and long-term follow-up. In general, pain, disability, and quality of life were significantly improved after an invasive intervention. Results after conservative treatments varied between significantly improved or unchanged. One study reported that patients had significantly worsened compared to baseline six years after following a rehabilitation program [109]. Since most studies reported significant improvement at follow-up, there was little difference in outcome at the different durations of follow-up.

Responder analyses

Setting aside the variety in definitions to determine clinical success and irrespective of the type of treatment patients received, we found that response on pain measures at long-term follow-up varied between 20 and 90% (10 studies with 15 treatment arms) and response on disability measures varied between 15 and 91% (22 studies with 32 treatment arms) (SDC 2).

Looking at different treatment types and taking into account the number of patients per treatment arm, clinical success on disability was achieved in 73% of patients that underwent a disc arthroplasty (n = 14 treatment arms), 75% of patients that underwent lumbar fusion (n = 7 treatment arms), 61% of patients that received multidisciplinary treatment or physiotherapy/exercise training (n = 4 treatment arms), and 63% of patients that received intradiscal therapies (n = 3 treatment arms). The only treatment type with > 3 treatments arms reporting response rates on pain measures was intradiscal therapy (n = 5 treatment arms), with 57% of patients achieving clinical success.

Discussion

The general purpose of this study was to identify and map the available evidence from long-term studies on chronic non-specific LBP. Our findings confirm the notion that there is little to no information available from natural cohorts when it comes to reporting on patient-centered outcomes other than pain. The majority (> 75%) of papers that were included examined long-term outcomes after invasive treatments. Surgical interventions, specifically lumbar fusion and disc arthroplasty, were most commonly reported. Among studies examining conservative treatments, physical therapy and multidisciplinary programs were most common. Overall, included studies were predominantly of moderate quality and differed in design, patient samples, and methods of data collection. These differences were most profound between studies on invasive and conservative treatments. In general, most studies reported improvements in pain and disability and, when measured, quality of life at long-term follow-up.

This review identifies several knowledge gaps regarding research into long-term outcomes of non-specific chronic LBP. First, there is still little insight into the natural course of LBP regarding outcomes such as disability, quality of
Table 2  Summary of reported results per treatment type and duration of follow-up (Non-specified for diagnosis or disease characteristics)

| Treatment arms (n) | Pain | Disability | Quality of life | Work | Health care utilization |
|-------------------|------|-----------|----------------|------|------------------------|
|                   | Outcome re-reported (n) | Effect | Outcome re-reported (n) | Effect | Outcome re-reported (n) | Effect | Outcome re-reported (n) | Effect | Outcome re-reported (n) | Effect |
|                   | +   | 0   | ?   | +   | 0   | ?   | +   | 0   | ?   | +   | 0   | ?   |
| Treatment type    |      |      |      |      |      |      |      |      |      |      |      |      |
| Invasive          |      |      |      |      |      |      |      |      |      |      |      |      |
| Lumbar fusion     | 41   | 34   |      | 23   | –   | –   | 11  | 38   |      | 26   | 1   | –   | 11  | 8   | 4   | 1   | –   | 3   | 10   | 1   | –   | 9   | nr  |
| Disc arthroplasty | 21   | 19   |      | 15   | –   | –   | 4   | 21   |      | 15   | 6   | –   | 8   | 5   | –   | –   | 3   | 2    | –   | –   | 2   | nr  |
| Intradiscal therapy | 7    | 7    |      | 6    | –   | –   | 1   | 7    |      | 6    | 1   | –   | 1   | 1   | –   | –   | 2   | –    | –   | –   | 2   | nr  |
| Implantable therapy | 3    | 3    |      | 3    | –   | –   | –   | 3    |      | 3    | –   | –   | –   | 2   | –   | –   | –   | 2    | –   | –   | 1   | nr  |
| Other             | 5    | 5    |      | 5    | –   | –   | –   | 5    |      | 5    | –   | –   | –   | 1   | –   | –   | –   | nr   | –   | –   | nr  |
| Conservative      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| Multidisciplinary treatment | 12   | 7    |      | 3    | 2   | –   | 2   | 10   |      | 5    | 2   | –   | 3   | 4   | –   | 1   | –   | 3   | 7    | –   | –   | 7   | 1    | –   | –   | –   |
| Physiotherapy or exercise training | 8    | 5    |      | –    | 1   | –   | 4   | 5    |      | –    | 1   | –   | 4   | nr  |      |      |      |      | 4    | –   | 1   | 3   | 1    | –   | 1   | –   |
| Cognitive treatment | 4    | 4    |      | –    | 1   | –   | 3   | 4    |      | –    | 1   | –   | 3   |      |      |      |      |      | 1    | –   | 1   | –   | 1    | –   | –   | –   |
| Usual (non-surgical) care | 4    | 2    |      | 1    | –   | –   | 1   | 2    |      | 1    | –   | –   | 1   | nr  |      |      |      |      | 1    | –   | –   | –   | 1    | nr  |
| Advice and/or education | 4    | 2    |      | 1    | –   | –   | 1   | 2    |      | 1    | –   | –   | 1   | nr  |      |      |      |      | 1    | –   | –   | –   | 1    | nr  |
| Other             | 5    | 5    |      | 2    | 1   | –   | 2   | 5    |      | 2    | 1   | 2   | 1    | –    | 1   | –   | –   | nr   | –   | 1   | –   | –   | nr  |
| No intervention   | 2    | 1    |      | 1    | –   | –   | –   | 1    |      | –    | 1   | –   | –   | nr  |      |      |      |      | 2    | –   | –   | –   | 2    | nr  |
| Sham/control for intradiscal therapy | 3    | 3    |      | –    | 1   | –   | 2   | 3    |      | –    | 1   | –   | 2   | nr  |      |      |      |      | nr   |      |      |      |      |
| Follow-up         |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 24–36 m           | 38   | 34   |      | 18   | 1   | –   | 15  | 36   |      | 22   | 1   | –   | 13  | 8    | 7   | 1   | –   | 6   | 2    | –   | 4   | 2    | 1    | 1   | –   | –   |
| 37–48 m           | 42   | 28   |      | 18   | 1   | –   | 9   | 32   |      | 20   | 3   | –   | 9   | 8    | 4   | 1   | –   | 3   | 8    | –   | 1   | 7    | nr  |
| 49–72 m           | 38   | 35   |      | 26   | 4   | –   | 5   | 35   |      | 26   | 3   | 5   | 12   | 8    | 1   | –   | 3   | 15   | –   | 1   | 14   | 1    | –   | 1   | –   |
| 73–216 m          | 33   | 29   |      | 22   | 2   | –   | 7   | 33   |      | 24   | 9   | –   | 8   | 5    | –   | 3   | 5    | –    | –   | 5   | 1    | –   | –   | 1    |

n, number of intervention groups; +, significantly improved (p < 0.05); 0, no significant change (p > 0.05); –, significantly worsened (p < 0.05); ?, no p value reported; nr, not reported; m, months

Under ‘Treatment type’, results of the latest follow-up measurement were reported. Under ‘Follow-up’, results for all follow-up measurements were reported.

One study reported results separately for ‘improved’ and ‘worsened/same’ subgroups of patients and was therefore not included in this table.
life, work, and health care utilization, because no natural cohorts met the inclusion criteria. In a natural cohort, subjects would be followed in real life in which numerous situations and interventions may appear. It is not limited to one or several specified interventions to study its effect. The studies included in this review examined clinical outcomes of non-specific LBP and concerned patients that were actively seeking health-care. Therefore, they might not be representative of people with sub-chronic or chronic LBP in the general population. Secondly, we noticed that repeated measurements during long-term follow-up were scarce. Only ten studies (11%) took more than one measurement after the two-year mark. These studies reported lasting improvements in symptoms after lumbar fusion [31, 32, 40, 41, 59, 72], disc arthroplasty [53, 58, 76, 92], and chiropractic care or primary care by an MD [102]. Nonetheless, recurrence of LBP is very common and studies with less than two years follow-up have also shown that post-treatment trajectories of pain and disability can vary a great deal between patients [122–124]. Third, the present review also affirms the notion that across LBP trials, the primary focus has been on pain and disability as outcome measures [125], even though other (generic) measures of health and well-being, such as quality of life and work (dis)ability have been recommended in core outcome sets to reflect the multidimensionality of LBP [15, 126–128]. Furthermore, few studies seem to monitor healthcare utilization during follow-up. These data can be challenging to collect; however, they are an important piece of the puzzle in determining whether outcomes at long-term follow-up might be the result of the original intervention (at baseline) or other interventions that were provide during follow-up. To conclude, in order to really understand both the (natural) course of LBP and results of LBP-related interventions over time, frequent measurements of relevant patient-centered outcomes are needed, as well as the use of complete core outcome sets including quality of life and work disability, and an overview of patients’ health care utilization during follow-up.

Even though the patient reported outcome measures in this review seem to reflect more positive long-term pain, disability and quality of life status compared to baseline measurements, this should not be misinterpreted as treatment effectiveness. This scoping review was not designed to study long-term effectiveness of interventions. A number of factors might have contributed to the appearance of consistent improvement years after experiencing persistent LBP. First, the reported improvements derive from statistical significance and do not necessarily imply clinical relevance. It is unclear whether patients perceived their improvement on different outcome measures as clinically relevant. Only a select number of studies performed a responder analysis. A previous review on outcome measures also reported that merely 8% of 401 included LBP trials reported a number or proportion of improved patients [125]. Although most of the studies in the present review that included a responder analysis reported high percentages of patients with clinically relevant improvement, cut-off scores for clinical success varied greatly. For instance, in some studies relative improvements of 25–30% on VAS or ODI were deemed successful, while others aimed for 50% [35–37, 95].

Other factors might also have influenced improvement in LBP symptoms. A previous review in patients with non-specific LBP found that response to primary care treatment followed a pattern of rapid early improvement followed by a plateau, regardless of whether active treatment, usual care, or placebo treatment was used [129]. Natural prognosis could be one explanation [10, 11, 130]. However, natural prognosis at long-term is mostly unknown. People are also more likely to seek health care at a time when their pain and symptoms are at their worst or most debilitating, which could further explain a positive overall course. Regression to the mean could also have played a role in the improvements in symptoms that were found after the start of treatment [131]. Overall, these factors likely influenced short-term improvements in LBP complaints, but if maintained, could also explain the reported long-term beneficial outcomes. Finally, publication and reporting bias cannot be ruled out. Only one study reported that patients had significantly worsened at long-term follow-up. Future (systematic) reviews on long-term studies on LBP should consider checking their findings against reported study protocols and/or unpublished trial data.

Surgical treatments are relatively over-represented in the present review. Safety issues and long-term adverse events are of more concern in surgical trials compared to conservative interventions, which may be why long-term data is collected and analyzed more often from invasive interventions. Also, surgical studies more often seem to utilize data that are retrospectively obtained from patient medical records [132, 133]. This makes it easier to collect and report long-term follow-up data. In spine surgery, complication incidence is potentially underestimated with retrospective assessments [134]; however, the present review includes results from PROMs and not occurrence of adverse events.

Studies on invasive and conservative treatments were notably different in their patient inclusion criteria. Invasive studies sought to include patients with disc-related diagnoses or symptoms, whereas conservative studies defined symptom-related criteria more generally (‘low back pain’). Although diagnoses based on lumbar structures (e.g., discogenic pain, facet joint pain) were very common in some settings, diagnostic tests do not reliably identify these structures as a source of LBP. The usefulness of these tests in clinical practice remains unclear [22, 26, 135] and current guidelines on LBP usually classify these diagnoses as non-specific [136]. Nevertheless, spine
surgeons have claimed that these diagnoses should classify as specific LBP and that better and earlier identification combined with, if indicated, invasive treatment would improve prognosis in these patients [137]. A Dutch task force that was tasked to develop a guideline for invasive treatment of lumbosacral pain syndromes has proposed to classify diagnoses such as facet joint pain, disc pain and FBSS as ‘degenerative uncomplicated spinal LBP syndromes’ [138]. In short, LBP diagnoses, as well as the decision to operate or treat conservatively, vary between countries and between medical disciplines. At present, there is no consensus among health care professionals on the classification of specific versus non-specific LBP. Improved consensus on a classification system could lead to more targeted care, reduce the need for expensive diagnostic methods, and facilitate comparison among LBP studies [17, 139, 140].

In line with worldwide research in the field of back pain, we identified a significant increase in annual publications on long-term outcomes of non-specific LBP [141]. The majority of selected studies were from Western countries, with the USA being the most productive (26% of studies). Little to no research took place in low- or middle-income countries, while in the past few decades the largest increases in disability due to LBP have occurred there [9, 142]. The impact of LBP in low- to middle-income countries potentially comes with disadvantages dissimilar to those in high-income countries and might therefore not be represented in the present review [9].

Finally, methodological quality of studies seemed to also increase over the years. Only prospectively conducted studies (prospective cohorts and RCT/CCTs) received a global ‘strong’ rating with the quality assessment tool that was utilized. Selection bias was often present in retrospectively conducted studies. In these instances, patients were included based on complete availability of follow-up data. Two sensitivity analyses were performed on the scoring method of the quality assessment tool. First, the global quality rating of a study was determined by the amount of ‘weak’ ratings that was scored on all separate domains. This means that studies that scored ‘moderate’ on each separate domain would have received a ‘strong’ global rating. A separate analysis showed that changing the global rating from strong to moderate for these studies would have had no effect on the results, since there were no studies that rated moderate on each domain. Second, prospective cohort studies received a ‘moderate’ rating on the domain study design. It could be argued that prospective cohort studies are a strong design for studying long-term outcomes. However, changing these ratings from moderate to strong on this domain would have also had no effect on the global quality rating.

Limitations

As to be expected, a number of studies on long-term LBP outcomes had to be excluded from this review after not meeting our inclusion criteria. This occurred most often with studies on samples with non-specific LBP mixed with specific LBP, samples with acute mixed with sub-acute and chronic LBP, and studies that failed to report baseline results of the outcomes measured at long-term follow-up. The latter in particular was common for measures related to health care utilization, since information has to be available, or recalled, from before baseline. Ultimately, only four studies could be included that reported health care use in the period before baseline [85, 99, 101, 114]. Another limitation is that this review gives limited insight into when the improvements that we observed took place. We chose to only report results from long-term follow-up (> 2 years), since the focus of was on mapping evidence from long-term follow-up studies. The complete course or trajectory of LBP symptoms could be studied in future reviews with a more narrow scope. Finally, the heterogeneity in the assessment and reporting of outcomes rendered it difficult to provide a qualitative synthesis of the results. A wide variety of instruments was used to measure pain, disability, quality of life, and work participation, and a considerable amount of studies did not report whether changes in scores between baseline and follow-up were statistically significant.

Conclusion

Patients with persistent non-specific LBP report improvements in pain, disability and quality of life years after seeking treatment. However, it remains unclear what factors might have influenced these improvements, and whether they are treatment-related. In part, because there is very little long-term evidence available from natural cohorts. Finally, studies that examined long-term outcomes of LBP symptoms varied greatly in design, quality, patient samples, and methods of data collection, and only few performed a responder analysis or applied repeated measurements after two years of follow-up.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00586-022-07111-3.

Author contributions All authors contributed to the study conception and design. The literature search was performed by AL Dutmer and data analysis was performed by AL Dutmer, R Soer, MF Reneman, and HR Schiphorst Preuper. The first draft of the manuscript was written by AL Dutmer and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.
meta-analysis. AJNR Am J Neuroradiol 36(12):2394–2399. https://doi.org/10.3174/ajnr.a4498
26. Maas ET, Juch JNS, Ostelo RWJG et al (2017) Systematic review of patient history and physical examination to diagnose chronic low back pain originating from the facet joints. Eur J Pain 21(3):403–414. https://doi.org/10.1002/ejp.963
27. Viera AJ, Garrett Joanne M (2005) Understanding interobserver agreement: the kappa statistic. Fam Med 37(5):360–363
28. Thomas BH, Ciliska D, Dobbins M et al (2004) A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evid Based Nurs 1(3):176–184. https://doi.org/10.1111/j.1524-475x.2004.00406.x
29. Armstrong-Stiles S, Stiles CR, Hagen NA et al (2012) Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract 18(1):12–18. https://doi.org/10.1111/j.1365-2753.2010.01516.x
31. Butterman GR, Hollmann S, Arpino J et al (2020) Value of sinusoplasty and cost-effectiveness analysis comparing posterior facet versus pedicle screw fixation. Eur Spine J 29(2):360–373. https://doi.org/10.1007/s00586-019-06165-0
33. Cousins S, Blencowe NS, Blazeby JM (2019) What is an invasive procedure? A definition to inform study design, evidence synthesis and research tracking. BMJ Open 9(7):E028576. https://doi.org/10.1136/bmjopen-2018-028576
35. Al-Kaisy A, Palmisani S, Smith TE et al (2018) Long-Term improvements in chronic axial low back pain patients without previous spinal surgery: a cohort analysis of 10-KHz high-frequency spinal cord stimulation over 36 months. Pain Med 19(6):1219–1226. https://doi.org/10.1093/pmj/pmx237
37. Formica C, Zanario A, Divano S et al (2020) Total disc replacement for lumbar degenerative disc disease: single centre 20 years experience. Spine J 29(7):1518–1526. https://doi.org/10.1016/j.spinee.2020.10.004
38. Geerdes BP, Geukers CW, van Erp WF (2001) Laparoscopic L5–S1 fusion versus open mini-ALIF, with a minimum 2-year follow-up. Eur Spine J 10(6):613–617. https://doi.org/10.1007/s00586-003-0526-y
40. Burkus JK, Gornet MF, Schuler TC et al (2009) Six-year outcomes of anterior lumbar interbody arthrodesis with use of interbody fusion cages and recombinant human bone morphogenetic protein-2. J Bone Joint Surg Am 91(5):1181–1189. https://doi.org/10.2106/jbjs.g.01485
41. Chung SK, Lee SH, Lim SR et al (2003) Comparative study of posterior facet versus pedicle screw fixation of circumferential fusion: a cohort study. Eur Spine J 23(2):347–355. https://doi.org/10.1007/s00586-013-2999-7
43. Cheng J, Zheng W, Wang H et al (2014) Posterolateral transfemoral selective endoscopic discectomy with thermal annuloplasty for discogenic low back pain: a prospective observational study. Spine (Phila Pa 1976) 39(26 Spec No):B60–B65. https://doi.org/10.1097/brs.0b013e31829f56f
45. Chung SK, Lee SH, Lim SR et al (2003) Comparative study of posterior facet versus pedicle screw fixation of circumferential fusion: a cohort study. Eur Spine J 23(2):347–355. https://doi.org/10.1007/s00586-006-0297-0
47. Di Silvestre M, Bakaloudis G, Lolli F et al (2009) Two-level total lumbar disc replacement. Eur Spine J 18(Suppl 1):62–70. https://doi.org/10.1007/s00586-009-0982-0
48. Fischgrund JS, Rhyne A, Macdaeg K et al (2020) Long-term outcomes following intraosseous basivertebreal nerve ablation for the treatment of chronic low back pain: 5-year treatment arm results from a prospective randomized double-blind sham-controlled multi-center study. Eur Spine J 29(8):1925–1934. https://doi.org/10.1007/s00586-020-06448-x
50. Gornet MF, Burkus JK, Dryer RF et al (2019) Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study. Neurosurg Spine 31(3):347–356. https://doi.org/10.3171/2019.2.spine181037
51. Gepstein R, Werner D, Shabat S et al (2005) Percutaneous posterior lumbar interbody fusion using the B-twin expandable spinal spacer. Minim Invasive Neurosurg 48(6):330–333
52. Giovita G, Mandelli D, Randelli F (2007) The Chariot III Artificial Disc lumbar disc prosthesis: assessment of medium-term results. J Orthop Traumatol 8(3):134–139. https://doi.org/10.1007/s00198-007-0122-7
53. Formica C, Zanario A, Divano S et al (2020) Total disc replacement for lumbar degenerative disc disease: single centre 20 years experience. Spine J 29(7):1518–1526. https://doi.org/10.1016/j.spinee.2020.10.004
55. Guyer RD, McAfée PC, Banco RJ et al (2009) Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion:
Katsimihas M, Baily CS, Issa K et al (2010) Prospective clinical and radiographic/neuroimaging documented outcome in transforminal lumbar interbody fusion. Neurosurg Focus 20(3):E8. https://doi.org/10.3171/foc.2006.20.3.9

Malham GM, Parker RM (2017) Early experience with lateral interbody fusion. Neurosurg Focus 20(3):E8. https://doi.org/10.2174/1876532701103010016

Madin SS, Boeree NN (2003) Circumferential and posterolateral fusion for lumbar disc disease. Clin Orthop Relat Res 409:114–123. https://doi.org/10.1097/01.blo.0000059581.08469.77

Maestretti G, Reischl N, Jacobi M et al (2011) Treatment of discogenic low back pain by total disc arthroplasty using the Prodisc prosthesis: analysis of a prospective cohort study with five-year clinical follow-up. Open Spine J 3:16–20. https://doi.org/10.2174/1876532710103010016

Malm GM, Parker RM (2017) Early experience with lateral lumbar total disc replacement: utility, complications and revision strategies. J Clin Neurosci 39:176–183. https://doi.org/10.1016/j.jocn.2017.01.033

Meir AR, Freeman BJC, Fraser RD et al (2013) Ten-year survival and clinical outcome of the AcroFlex lumbar disc replacement for the treatment of symptomatic disc degeneration. Spine J 13(1):13–21. https://doi.org/10.1016/j.spinee.2012.12.008

Niemeyer T, Bövingloh AS, Halm H et al (2004) Results after anterior–posterior lumbar spinal fusion: 2–5 years follow-up. Int Orthop 28(5):298–302. https://doi.org/10.1007/s00264-004-0577-7

Noreiga DC, Ardura F, Hernández-Ramojo R et al (2021) Treatment of degenerative disc disease with allogeneic mesenchymal stem cells: long-term follow-up results. Transplantation 105(2):E25–E27. https://doi.org/10.1097/tp.0000000000003471

Nunley PD, Jawahar A, Brandao SM et al (2008) Intradiscal electrothermal therapy (IDET) for low back pain in worker’s compensation patients: can it provide a potential answer? Long-term results. J Spinal Disord Tech 21(1):11–18. https://doi.org/10.1097/bsd.0b013e31840c990e

Nyström B, Weber H, Schillberg B et al (2018) Symptoms and signs possibly indicating segmental, discogenic pain. A fusion study with 18 years of follow-up. Scand J Pain 16:213–220. https://doi.org/10.1016/j.sjpain.2016.10.007

Ohtori S, Kinoshita T, Yamashita M et al (2009) Results of surgery for discogenic low back pain: a randomized study using discography versus discoblock for diagnosis. Spine (Phila Pa 1976) 34(13):1345–8. https://doi.org/10.1097/brs.0b013e3181 a401bf

Pan F, Shen B, Chy SK et al (2016) Transforminal endoscopic system technique for discogenic low back pain: a prospective cohort study. Int J Surg 35:134–138. https://doi.org/10.1016/j. ijsu.2016.09.091

Park C, Ryu K, Lee K et al (2012) Clinical outcome of lumbar total disc replacement using ProDisc-L in degenerative disc disease: minimum 5-year follow-up results at a single institute. Spine (Phila Pa 1976) 37(8):672–7. https://doi.org/10.1097/ bhrs.0b013e3182 2ed85

Park S, Lee C, Chang S et al (2016) Long-term outcomes following lumbar total disc replacement using ProDisc-II: average 10-year follow-up at a single institute. Spine (Phila Pa 1976) 41(11):971–7. https://doi.org/10.1097/ bhrs.0b013e31822ed99

Peng B, Chen J, Kuang Z et al (2009) Diagnosis and surgical treatment of back pain originating from endplate. Eur Spine J 18(7):1035–1040. https://doi.org/10.1007/ s00586-009-0938-4

Petilon J, Roth J, Hardenbrook M (2011) Results of lumbar total disc arthroplasty in military personnel. J Spinal Disord Tech 24(5):297–301. https://doi.org/10.1097/ bsd.0b013e3181bf 3e2a

Pettine K, Suzuki R, Sand T et al (2017) Autologous bone marrow concentrate intradiscal injection for the treatment of degenerative disc disease with three-year follow-up. Int Orthop 41(10):2097–2103. https://doi.org/10.1007/s00264-017-3560-9

Pihlajamäki H, Böstman O, Ruuskanen M et al (1996) Postero-lateral lumbar fusion with transpedicular fixation: 63 consecutive cases followed for 4 (2–6) years. Acta Orthop Scand 67(1):63–68. https://doi.org/10.3109/1745367960895612

Pimenta L, Marchi L, Oliveira L et al (2013) A prospective, randomized, controlled trial comparing radiographic and clinical outcomes between stand-alone lateral interbody lumbar fusion with either silicate calcium phosphate or rh-BMP2. J Neurourol Surg A Cent Eur Neurosurg 74(6):343–350. https://doi.org/10.5555/ s0032-1333420

Pimenta L, Marchi L, Oliveira L et al (2013) A prospective, randomized, controlled trial comparing radiographic and clinical outcomes between stand-alone lateral interbody lumbar fusion with either silicate calcium phosphate or rh-BMP2. J Neurourol Surg A Cent Eur Neurosurg 74(6):343–350. https://doi.org/10.5555/ s0032-1333420

Pimenta L, Marchi L, Oliveira L et al (2018) Elasticomeric lumbar total disc replacement: clinical and radiological results with minimum 84 months follow-up. Int J Spine Surg 12(1):49–57. https://doi.org/10.14444/5009

Plais N, Theyenot X, Cogniet A et al (2018) Maverick total disc arthroplasty performs well at 10 years follow-up: a prospective study with HRQL and balance analysis. Eur Spine J 27(3):720–727. https://doi.org/10.1007/s00586-017-5065-z

Pokorny G, Marchi L, Amaral R et al (2019) Lumbar total disc replacement by the lateral approach–up to 10 years follow-up. World Neurosurg 122:e325–e333. https://doi.org/10.1016/j. wneu.2018.10.033

Putzier M, Hoff E, Tohtz S et al (2010) Dynamic stabilization adjacent to single-level fusion: part II. No clinical benefit for asymptomatic, initially degenerated adjacent segments after 6
years follow-up. Eur Spine J 19(12):2181–9. https://doi.org/10.1007/s00586-010-1517-4
86. Raphael JH, Southall JL, Gnanadurai TV et al (2002) Long-term experience with implanted intrathecal drug administration systems for failed back syndrome and chronic mechanical low back pain. BMC Musculoskeletal Disord 3:17. https://doi.org/10.1186/1471-2474-3-17
87. Ren D, Liu X, Du S et al (2015) Percutaneous nucleoplasty using coaglation technique for the treatment of chronic nonspecific low back pain: 5-year follow-up results. Chin Med J (Engl) 128(14):1893–1897. https://doi.org/10.4103/0366-6999.160518
88. Rouben D, Casnellie M, Ferguson M (2011) Long-term durability of minimal invasive posterior transforaminal lumbar interbody fusion: a clinical and radiographic follow-up. J Spinal Disord Tech 24(5):288–296. https://doi.org/10.1097/bsd.0b013e31819d60a
89. Saal JA, Saal JS (2002) Intradiscal electrothermal treatment for chronic discogenic low back pain: prospective outcome study with a minimum 2-year follow-up. Spine (Phila Pa 1976) 27(9):966–73. https://doi.org/10.1097/00007632-200205010-00017
90. Schimmel JJP, Poelshmann MS, Horsting PP et al (2016) PEEK cages in lumbar fusion: mid-term clinical outcome and radiologic fusion. Clin Spine Surg 29(5):E252–E258. https://doi.org/10.1097/bss.0b013e3182eeaf74
91. Schulte TL, Leistra F, Bullmann V et al (2007) Disc height reduction in adjacent segments and clinical outcome 10 years after lumbar 360 degrees fusion. Eur Spine J 16(12):2152–2158. https://doi.org/10.1007/s00586-007-0515-7
92. Seipe CI, Heider F,Wiechert K et al (2014) Mid- to long-term results of total lumbar disc replacement: a prospective analysis with 5- to 10-year follow-up. Spine J 14(8):1417–1431. https://doi.org/10.1016/j.spinee.2013.08.028
93. Sköld C, Tropp H, Berg S (2013) Five-year follow-up of total disc replacement compared to fusion: a randomized controlled trial. Eur Spine J 22(10):2288–2295. https://doi.org/10.1007/s00586-013-2926-y
94. Strube P, Hoff EK, Schürings M et al (2013) Motion-preservation following total lumbar disc replacement at the lumbosacral junction: Part 2: distraction and posterior translation lead to clinical failure after a mean follow-up of 5 years. Eur Spine J 22(10):2279–87. https://doi.org/10.1007/s00586-013-2967-2
95. Thalgott JS, Kiezl Z, Timlin M et al (2002) Anterior lumbar interbody fusion with processed sea coral (coralline hydroxypatite) as part of a circumferential fusion. Spine (Phila Pa 1976) 27(24):E518–E25. https://doi.org/10.1097/00007632-200212150-00011
96. Wuertinger C, Annes RDA, Hitzl W et al (2018) Motion preservation following total lumbar disc replacement at the lumbosacral junction: a prospective long-term clinical and radiographic investigation. Spine J 18(1):72–80. https://doi.org/10.1016/j.spinee.2017.06.035
97. Zielstra DJ, Staartjes VE, Schröder ML (2017) Minimally invasive translumbar lumbosacral interbody fusion: a ten year single-centre experience. Int Orthop 41(1):113–119
98. Bendix AE, Bendix T, Haestrup C et al (1998) A prospective, randomized 5-year follow-up study of functional restoration in chronic low back pain patients. Eur Spine J 7(2):111–119. https://doi.org/10.1007/s005860050040
99. Bentsen H, Lindgård F, Manthorpe R (1997) The effect of dynamic strength back exercise and/or a home training program in 57-year-old women with chronic low back pain. Results of a prospective randomized study with a 3-year follow-up period. Spine (Phila Pa 1976) 22(13):1494–500. https://doi.org/10.1097/00007632-199707100-00014
100. Carvalho C, Pais M, Cunha L et al (2021) Open-label placebo for chronic low back pain: a 5-year follow-up. Pain 162(5):1521–1527. https://doi.org/10.1007/j.pain.000000000002162
101. Groot D, van Hooft ML, Kroeeze R et al (2019) Long-term results of an intensive cognitive behavioral pain management program for patients with chronic low back pain: a concise report of an extended cohort with a minimum of 5-year follow-up. Eur Spine J 28(7):1579–1585. https://doi.org/10.1007/s00586-019-05967-6
102. Haas M, Goldberg B, Aickin M et al (2004) A practice-based study of patients with acute and chronic low back pain attending primary care and chiropractic physicians: two-week to 48-month follow-up. J Manip Physiol Ther 27(3):160–169. https://doi.org/10.1016/j.mpjt.2003.12.020
103. Hamre HJ, Kiene H, Glockmann A et al (2013) Long-term outcomes of anthroposophic treatment for chronic disease: a four-year follow-up analysis of 1510 patients from a prospective observational study in routine outpatient settings. BMC Res Notes 6:269. https://doi.org/10.1186/1756-0500-6-269
104. Lamb SE, Mistry D, Lall R et al (2012) Group cognitive behavioural interventions for low back pain in primary care: extended follow-up of the Back Skills Training Trial (ISRCTN54717854). Pain 153(2):499–501. https://doi.org/10.1016/j.pain.2011.11.016
105. Lanes TC, Gauron EF, Sprat KF et al (1995) Long-term follow-up of patients with chronic back pain treated in a multidisciplinary rehabilitation program. Spine (Phila Pa 1976) 20(7):801–6. https://doi.org/10.1097/00007632-199504000-00012
106. Lankhorst GJ, van der Stadt RJ, van der Korst JK (1985) The natural history of idiopathic low back pain. A three-year follow-up study of spinal motion, pain and functional capacity. Scand J Rehabil Med 17(1):1–4
107. Patrick LE, Altnaier EM, Found EM (2004) Long-term outcomes in multidisciplinary treatment of chronic low back pain: results of a 13-year follow-up study. Spine (Phila Pa 1976) 29(8):850–5. https://doi.org/10.1097/00007632-200404150-00006
108. Peng B, Fu X, Pang X et al (2012) Prospective clinical study on natural history of discogenic low back pain at 4 years of follow-up. Pain Phys 15(6):525–532
109. Raak R, Wikblad K, Raak A Sr et al (2002) Catastrophizing and health-related quality of life: a 6-year follow-up of patients with chronic low back pain. Rehabil Nurs 27(3):110–116. https://doi.org/10.1016/j.rehab.2004.07.009.e1
110. Rantzen J, Karppinen YA et al (2018) Effectiveness of three interventions for secondary prevention of low back pain in the occupational health setting—a randomised controlled trial with a natural course control. BMC Public Health 18(1):598. https://doi.org/10.1186/s12889-018-5476-8
111. Rasmussen-Barr E, Ang B, Arvidsson L et al (2009) Graded exercise for recurrent low-back pain: a randomized, controlled trial with 6-, 12-, and 36-month follow-ups. Spine (Phila Pa 1976) 34(3):221–8. https://doi.org/10.1097/0b013e318138191e7cb
112. Rhyne AL, Smith SE, Wood KE et al (1995) Outcome of unoperated discogram-positive low back pain. Spine (Phila Pa 1976) 20(18):1997–2000. https://doi.org/10.1097/00007632-199509150-00007
113. Ulby PM, Bendix T, Ohrt-Nissen S et al (2019) Modic changes are not associated with long-term pain and disability: a cohort study with 13-year follow-up. Spine (Phila Pa 1976) 44(17):1186–92. https://doi.org/10.1097/BRS.0000000000003051
114. Van Hoof W, O’Sullivan K, Verschueren S et al (2021) Evaluation of absenteeism, pain, and disability in nurses with persistent low back pain following cognitive functional therapy: a case series pilot study with 3-year follow-up. Phys Ther 101(1):pzaa164. https://doi.org/10.1093/ptj/pzaa164
115. Vibe Fersum K, Smith A, Kvale A et al (2019) Cognitive functional therapy in patients with non-specific chronic low back
paign—a randomized controlled trial 3-year follow-up. Eur J Pain 23(8):1416–1424. https://doi.org/10.1002/ejp.1399
116. Brox JI, Nygaard OP, Holm I et al (2010) Four-year follow-up of surgical versus non-surgical therapy for chronic low back pain. Ann Rheum Dis 69(9):1643–1648. https://doi.org/10.1136/ard.2009.108902
117. Froholdt A, Holm I, Keller A et al (2011) No difference in long-term trunk muscle strength, cross-sectional area, and density in patients with chronic low back pain 7 to 11 years after lumbar fusion versus cognitive intervention and exercises. Spine J 11(8):718–725. https://doi.org/10.1016/j.spinee.2011.06.004
118. Froholdt A, Reikeraas O, Holm I et al (2012) No difference in 9-year outcome in CLBP patients randomized to lumbar fusion versus cognitive intervention and exercises. Eur Spine J 21(12):2531–2538. https://doi.org/10.1007/s00586-012-2382-0
119. Furunes H, Storheim K, Brox JI et al (2017) Total disc replacement versus multidisciplinary rehabilitation in patients with chronic low back pain and degenerative discs: 8-year follow-up of a randomized controlled multicenter trial. Spine J 17(10):1480–1488. https://doi.org/10.1016/j.spinee.2017.05.011
120. Hedlund R, Johansson C, Hägg O et al (2016) The long-term outcome of lumbar fusion in the Swedish lumbar spine study. Spine J 16(5):579–587. https://doi.org/10.1016/j.spinee.2015.08.065
121. Kleimeyer JP, Cheng J, Alamin TF et al (2018) Selective anterior lumbar interbody fusion for low back pain associated with degenerative disc disease versus nonsurgical management. Spine (Phila Pa 1976) 43(19):1372–80. https://doi.org/10.1097/brs.0000000000002630
122. Åxén I, Leboeuf-Yde C (2013) Trajectories of low back pain. Best Pract Res Clin Rheumatol 27(5):601–612. https://doi.org/10.1016/j.berh.2013.10.004
123. Kongsted A, Kent P, Åxén I (2016) What have we learned from ten years of trajectory research in low back pain? BMC Musculoskeletal Disord 17:220. https://doi.org/10.1186/s12891-016-1071-2
124. Dutmer AL, Schiphorst Preuper HR, Stewart RE et al (2020) Trajectories of disability and low back pain impact: 2-year follow-up of the groningen spine cohort. Spine (Phila Pa 1976) 45(23):1649–60. https://doi.org/10.1097/brs.0000000000003647
125. Froud R, Patel S, Rajendran D et al (2016) A systematic review of outcome measures use, analytical approaches, reporting methods, and publication volume by year in low back pain trials published between 1980 and 2012: respice, adspice, et prospice. Pain Pract 16(1):90–110. https://doi.org/10.1111/papr.012318
126. Deyo RA, Battie M, Beurskens AJ et al (1998) Outcome measures for low back pain research. A proposal for standardized use. Spine (Phila Pa 1976) 23(18):2003–13. https://doi.org/10.1097/00007632-199809150-00018
127. Froud R, Ellard D, Patel S et al (2015) Primary outcome measure use in back pain trials may need radical reassessment. BMC Musculoskeletal Disord 16:88. https://doi.org/10.1186/s12891-015-0534-1
128. Chiarotto A, Boers M, Deyo RA et al (2018) Core outcome measurement instruments for clinical trials in nonspecific low back pain. Pain 159(3):481–495. https://doi.org/10.1097/j.pain.0000000000001117
129. Artus M, van der Windt DA, Jordan KP et al (2010) Low back pain symptoms show a similar pattern of improvement following a wide range of primary care treatments: a systematic review of randomized clinical trials. Rheumatology (Oxford) 49(12):2346–2356. https://doi.org/10.1093/rheumatology/keq245
130. Pengel LHM, Herbert RD, Maher CG et al (2003) Acute low back pain: systematic review of its prognosis. BMJ 327(7410):323. https://doi.org/10.1136/bmj.327.7410.323
131. Morton V, Torgerson DJ (2005) Regression to the mean: treatment effect without the intervention. J Eval Clin Pract 11(1):59–65. https://doi.org/10.1111/j.1365-2753.2004.00505.x
132. Sauerland S, Lefering R, Neugebauer EAM (2002) Retrospective clinical studies in surgery: potentials and pitfalls. J Hand Surg Br 27(2):117–121. https://doi.org/10.1054/jhsb.2001.0703
133. Hanson B, Kopjar B (2005) Clinical studies in spinal surgery. Eur Spine J 14(8):721–725. https://doi.org/10.1007/s00586-005-0926-2
134. Campbell PG, Malone J, Yadla S et al (2011) Comparison of ICD-9-based, retrospective, and prospective assessments of perioperative complications: assessment of accuracy in reporting. J Neurosurg Spine 14(1):16–22. https://doi.org/10.3171/2010.9.spine10151
135. Willems PC, Staal JB, Walenkamp GHIM et al (2013) Spinal fusion for chronic low back pain: systematic review on the accuracy of tests for patient selection. Spine J 13(2):99–109. https://doi.org/10.1016/j.spinee.2012.10.001
136. Koes BW, van Tulder MW, Ostelo R et al (2001) Clinical guidelines for the management of low back pain in primary care: an international comparison. Spine (Phila Pa 1976) 26(22):2504–13. https://doi.org/10.1097/00007632-200111150-00022
137. Willems P, de Bie R, Oner C et al (2011) Clinical decision making in spinal fusion for chronic low back pain. Results of a nationwide survey among spine surgeons. BMJ Open 1(2):e000391. https://doi.org/10.1136/bmjopen-2011-000391
138. Itz CJ, Willems PC, Zielstra DJ et al (2016) Dutch multidisciplinary guideline for invasive treatment of pain syndromes of the lumbosacral spine. Pain Pract 16(1):90–110. https://doi.org/10.1111/papr.12318
139. Dewitte V, de Pauw R, De Meulemeester R et al (2018) Classical classification criteria for nonspecific low back pain: a Delphi-survey of clinical experts. Musculoskelet Sci Pract 34:66–76. https://doi.org/10.1016/j.msyp.2018.01.002
140. Amundsen PA, Evans DW, Rajendran D et al (2018) Inclusion and exclusion criteria used in non-specific low back pain trials: a review of randomised controlled trials published between 2006 and 2012. BMC Musculoskeletal Disord 19(1):113. https://doi.org/10.1186/s12891-018-2034-6
141. Wang R, Weng L, Peng M et al (2020) Exercise for low back pain: a bibliometric analysis of global research from 1980 to 2018. J Rehabil Med 52(4):jrm00052. https://doi.org/10.2340/16501977-2674
142. Hoy DG, Smith E, Cross M et al (2015) Reflecting on the global burden of musculoskeletal conditions: lessons learnt from the global burden of disease 2010 study and the next steps forward. Ann Rheum Dis 74(1):4–7. https://doi.org/10.1136/annrheumdis-2014-205393

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.