Cochrane Library

Cochrane Database of Systematic Reviews

Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor review (Review)

Hayden JA, Wilson MN, Riley RD, Iles R, Pincus T, Ogilvie R

Hayden JA, Wilson MN, Riley RD, Iles R, Pincus T, Ogilvie R.
Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor review.
Cochrane Database of Systematic Reviews 2019, Issue 11. Art. No.: CD011284.
DOI: 10.1002/14651858.CD011284.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Prognosis Review]

Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor review

Jill A Hayden[1], Maria N Wilson[2], Richard D Riley[3], Ross Iles[4], Tamar Pincus[5], Rachel Ogilvie[6]

[1]Department of Community Health & Epidemiology, Dalhousie University, Halifax, Canada. [2]Department of Community Health and Epidemiology, Dalhousie University, Halifax, Canada. [3]School of Primary, Community and Social Care, Keele University, Keele, UK. [4]Department of Physiotherapy, Faculty of Medicine, Nursing and Health Sciences, Monash University, Frankston, Australia. [5]Department of Psychology, Royal Holloway University of London, Egham, UK. [6]Community Health & Epidemiology, Dalhousie University, Halifax, Canada

Contact address: Jill A Hayden, Department of Community Health & Epidemiology, Dalhousie University, 5790 University Avenue, Room 403, Halifax, NS, B3H 1V7, Canada. jhayden@dal.ca.

Editorial group: Cochrane Back and Neck Group.
Publication status and date: New, published in Issue 11, 2019.

Citation: Hayden JA, Wilson MN, Riley RD, Iles R, Pincus T, Ogilvie R. Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor review. *Cochrane Database of Systematic Reviews* 2019, Issue 11. Art. No.: CD011284. DOI: 10.1002/14651858.CD011284.pub2.

ABSTRACT

Background

Low back pain is costly and disabling. Prognostic factor evidence can help healthcare providers and patients understand likely prognosis, inform the development of prediction models to identify subgroups, and may inform new treatment strategies. Recent studies have suggested that people who have poor expectations for recovery experience more back pain disability, but study results have differed.

Objectives

To synthesise evidence on the association between recovery expectations and disability outcomes in adults with low back pain, and explore sources of heterogeneity.

Search methods

The search strategy included broad and focused electronic searches of MEDLINE, Embase, CINAHL, and PsycINFO to 12 March 2019, reference list searches of relevant reviews and included studies, and citation searches of relevant expectation measurement tools.

Selection criteria

We included low back pain prognosis studies from any setting assessing general, self-efficacy, and treatment expectations (measured dichotomously and continuously on a 0 - 10 scale), and their association with work participation, clinically important recovery, functional limitations, or pain intensity outcomes at short (3 months), medium (6 months), long (12 months), and very long (> 16 months) follow-up.

Data collection and analysis

We extracted study characteristics and all reported estimates of unadjusted and adjusted associations between expectations and related outcomes. Two review authors independently assessed risks of bias using the Quality in Prognosis Studies (QUIPS) tool. We conducted narrative syntheses and meta-analyses when appropriate unadjusted or adjusted estimates were available. Two review authors independently graded and reported the overall quality of evidence.

Main results

We screened 4635 unique citations to include 60 studies (30,530 participants). Thirty-five studies were conducted in Europe, 21 in North America, and four in Australia. Study populations were mostly chronic (37%), from healthcare (62%) or occupational settings (26%). General expectation was the most common type of recovery expectation measured (70%); 16 studies measured more than one type of expectation.

Usable data for syntheses were available for 52 studies (87% of studies; 28,885 participants). We found moderate-quality evidence that positive recovery expectations are strongly associated with better work participation (narrative synthesis: 21 studies; meta-analysis: 12 studies, 4777 participants: odds ratio (OR) 2.43, 95% confidence interval (CI) 1.64 to 3.62), and low-quality evidence for clinically important recovery outcomes (narrative synthesis: 12 studies; meta-analysis: 5 studies, 1820 participants: OR 1.89, 95% CI 1.49 to 2.41), both at follow-up times closest to 12 months, using adjusted data. The association of recovery expectations with other outcomes of interest, including functional limitations (narrative synthesis: 10 studies; meta-analysis: 3 studies, 1435 participants: OR 1.40, 95% CI 0.85 to 2.31) and pain intensity (narrative synthesis: 9 studies; meta-analysis: 3 studies, 1555 participants: OR 1.15, 95% CI 1.08 to 1.23) outcomes at follow-up times closest to 12 months using adjusted data, is less certain, achieving very low- and low-quality evidence, respectively. No studies reported statistically significant or clinically important negative associations between recovery expectations and any low back pain outcome.

Authors' conclusions

We found that individual recovery expectations are probably strongly associated with future work participation (moderate-quality evidence) and may be associated with clinically important recovery outcomes (low-quality evidence). The association of recovery expectations with other outcomes of interest is less certain. Our findings suggest that recovery expectations should be considered in future studies, to improve prognosis and management of low back pain.

PLAIN LANGUAGE SUMMARY

The impact of individual recovery expectations on pain, limitations in activities and return to work in low back pain

What is the aim of this review?

The aim of this Cochrane Review is to find out if positive recovery expectations of people with low back pain are related to their future pain, activities they are able to do and return to work. Are people who think they will recover from their low back pain more likely to get better?

Key messages

People with low back pain who have positive expectations of their own recovery are more likely to return to work and to recover from pain and increase the activities they are able to do.

What was studied in this review?

Low back pain is costly and causes a lot of disability. It is important to understand what characteristics of a person with low back pain are connected with how well they will recover (also known as their 'prognosis'). People's characteristics are often not changeable, including a characteristic like age. However, there is evidence that someone's expectations of recovery may be changeable. If positive expectations are indeed connected to improved back pain outcomes then helping a person to have positive expectations of their own recovery may help them to recover.

For this review, we examined three types of recovery expectations and their relation to back pain outcomes: general expectations of recovery (e.g. will your back pain last only a short time?), self-efficacy expectations (e.g. do you believe you will be able to return to your normal activities?) and treatment expectations (e.g. will physiotherapy improve your back pain?).

What are the main results of this review?

We reviewed 4635 references and included 60 relevant studies. These studies included information about 30,530 people with low back pain. They looked at people's expectations of their own recovery and how that was related to their pain, limitations in activities and return to work one year after their back pain episode.

Overall, we found good evidence that positive expectations of recovery are related to a higher likelihood of returning to work. The evidence about positive recovery expectations with other recovery, limitations in activities and pain intensity outcomes is not as strong. We did not find any studies that showed that positive expectations of recovery were related to worse low back pain outcomes.

How up-to-date is this review?

The review authors searched for studies that had been published up to 12 March 2019.

Summary of findings for the main comparison.

Individual recovery expectations as a prognostic factor for low back pain

Patient or population: People with non-specific low back pain presenting to healthcare, occupational, general or mixed populations

Prognostic factor: Individual recovery expectations (measured dichotomously or continuously, as noted)

Outcomes	Reported adjusted associations, # studies (# participants)	Phase of investigation, # studies exploratory; confirmatory (# participants)	Meta-analysis, # study groups (# participants)	Meta-analysis relative effect (95% CI)	Quality of the evidence (GRADE)	Comments (rating of factors considered)
WORK PARTICIPATION Follow-up: closest to 12 months	Positive: 16 (4324) Neutral: 5 (2473) Negative: 0	16 E (5529); 5 C (1268)	13 (4777)[a]	**OR 2.43** (1.64 to 3.62)[a]	⊕⊕⊕⊖ **moderate**	**Limitations (ROB):** Serious **Inconsistency:** No serious **Indirectness:** No serious **Imprecision:** No serious **Publication/reporting bias:** Serious **Effect size reported:** Mod-Large **Dose effect:** N/A **Confirmatory evidence:** Available
IMPORTANT RECOVERY Follow-up: closest to 12 months	Positive: 6 (7265) Neutral: 6 (996) Negative: 0	12 E (8261); 0 C	5 (1820)[a]	**OR 1.89** (1.49 to 2.41)[a]	⊕⊕⊖⊖ **low**	**Limitations (ROB):** Serious **Inconsistency:** No serious **Indirectness:** No serious **Imprecision:** No serious **Publication/reporting bias:** Serious **Effect size reported:** Mod-Large

						Dose effect: N/A **Confirmatory evidence:** Not available
FUNCTIONAL LIMITATIONS Follow-up: closest to 12 months	Positive: 6 (1397) Neutral: 4 (2079) Negative: 0	6 E (2825); 4 C (651)	3 (1435)[b]	**OR 1.40** (0.85 to 2.31)[b]	⊕⊖⊖⊖ **very low**	**Limitations (ROB):** Serious **Inconsistency:** Serious **Indirectness:** No serious **Imprecision:** Serious **Publication/reporting bias:** Serious **Effect size reported:** Small **Dose effect:** N/A **Confirmatory evidence:** Available
PAIN INTENSITY Follow up: closest to 12 months	Positive: 5 (1510) Neutral: 4 (1216) Negative: 0	4 E (1174); 5 C (1552)	3 (1555)[b]	**OR 1.15** (1.08 to 1.23)[b]	⊕⊕⊖⊖ **low**	**Limitations (ROB):** Serious **Inconsistency:** No serious **Indirectness:** No serious **Imprecision:** No serious **Publication/reporting bias:** Serious **Effect size reported:** Small **Dose effect:** N/A **Confirmatory evidence:** Available

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

C = confirmatory phase study; E = exploratory phase study; OR = odds ratio; ROB = risk of bias

[a]Dichotomous measure of expectations (adjusted results; follow-up closest to 12 months).
[b]Continuous measure of expectations (scale /10; adjusted results; follow-up closest to 12 months).

BACKGROUND

Description of the health condition and context

Low back pain is one of the most common health conditions, and has high socioeconomic impact (Freburger 2009; Hoy 2010; Lim 2012). Approximately 540 million people are estimated to have nonspecific low back pain (GBD Collaborators 2016; Hartvigsen 2018b), and low back pain was identified in the most recent Global Burden of Disease study as the leading cause of disability globally (GBD Collaborators 2016). There is evidence that the prevalence and associated costs of low back pain are rising (Freburger 2009).

Researchers define low back pain as pain on the posterior aspect of the body from the lower margin of the twelfth ribs to the lower gluteal folds, with or without pain referred to the leg(s), that is severe enough to limit usual activities for more than one day (Dionne 2008). Most people who experience low back pain have 'non-specific low back pain', a diagnosis of exclusion that includes heterogeneous presentation and symptoms not attributed to a recognisable, known specific pathology (for example, fracture, rheumatoid arthritis, infection, neoplasm, or metastasis).

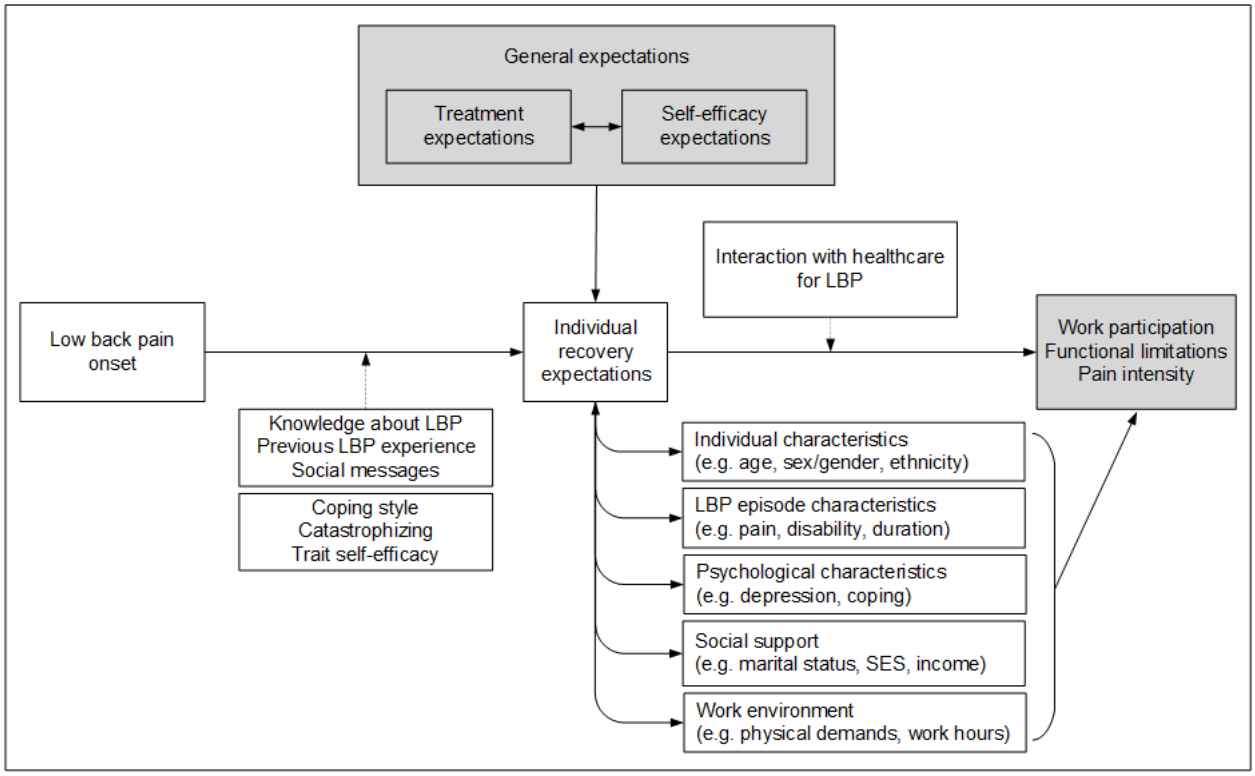
Most the social and economic costs associated with low back pain are attributed to a small number of sufferers who have prolonged disability and require increased use of health services and time off work (Freburger 2009; Hayden 2010). Most individuals experiencing a new episode of low back pain will recover within a few weeks. However, a quarter to a third will continue to report low back pain after 12 months (Hayden 2010). Recurrences are common and individuals who develop chronic, longstanding low back pain tend to show a more persistent course (Hayden 2010); studies of chronic low back pain indicate that 42% to 75% from general populations (Hestbaek 2003), and 60% to 80% from healthcare consulting populations (Hayden 2010) will continue to have low back pain after one year.

Consideration of prognosis and prognostic factors are important in low back pain research and treatment. It has not been possible to identify a specific cause for most cases of low back pain, and interventions with strong evidence of effectiveness have not been identified. Research studies have found many factors to be associated with a poor outcome in low back pain, often with conflicting results (Hayden 2007). A 'review of reviews' study found that several factors were consistently reported to be associated with a poor outcome, including individual characteristics (older age, poor general health), factors related to the back pain episode characteristics (baseline disability, sciatica), and psychological characteristics (increased stress, negative cognitive characteristics), as well as social supports and the work environment (poor relations with colleagues, heavy physical demands, receipt of compensation) (Hayden 2009). However, there is still substantial inconsistency in findings reported across low back pain studies. There is also a need for more attention to be paid to the collection and consideration of prognostic factors within research studies, which could include considering prognostic factors in trial randomisation strategies, or adjusting for these factors in analyses. High-quality evidence about prognostic factors associated with outcomes can improve management of low back pain by helping healthcare providers and patients to understand the likely prognosis, and can inform other research. This could include the development/refinement of outcome prediction models to identify subgroups of people with low back pain, identification of treatment effect modifiers, and influencing the development of new treatment strategies considering modifiable prognostic factors that potentially cause poor outcomes (Riley 2013).

Description of the prognostic factor

This Cochrane Review explores individual recovery expectations, a potentially modifiable prognostic factor that has shown promise in existing low back pain prognostic factor reviews (Fadyl 2008; Iles 2008; Iles 2009). Recovery expectations are what the individual 'expects will occur' in the future from their health condition. We referred to the Social Cognitive Theory (Bandura 1977; Bandura 2004) to develop a theoretical framework that guided our assessment of evidence about individual recovery expectations. In this model, individual recovery expectations involve cognitive processing and may be informed by past personal experience, knowledge and beliefs, and suggestions from or observations of other people. We consider three types of related individual recovery expectations relevant to the low back pain field: general expectations, self-efficacy expectations, and treatment expectations. General expectations are broadly-defined recovery expectations, related to a future low back pain outcome; an example of a single-item question is: "I expect to return to work within six months", or "My low back pain will last a short time". Self-efficacy expectations are a person's perceptions about their ability to execute behaviours to achieve a future outcome; for example: "I believe that I will be able to do my usual work activities to return to my job", or "I am confident that I will be able to learn to cope with the pain and get back to my normal activities". Treatment expectations are expectations of future low back pain outcomes specifically related to ongoing treatment; for example: "My treatment will help improve my low back pain", or "My treatment can prevent my back pain from getting worse". Figure 1 presents our conceptual framework of the relationship between individual recovery expectations (hereafter referred to as 'expectations') and low back pain outcomes.

Figure 1. Conceptual framework of the relationship between recovery expectations and LBP outcomes.



Health outcomes

Expectations may be related to low back pain outcomes through several possible pathways. These include modifying individual coping behaviours, withdrawal related to fear of pain or low mood, or by influencing treatment compliance or seeking health care. In Social Cognitive Theory, Bandura proposed that self-efficacy expectations can modify individual behaviours by determining the amount of effort that a person will exert to cope with their health condition (Bandura 1977; Bandura 2004). Following the fear-avoidance model (Vlaeyen 2000), processes related to the fear of pain may lead to avoiding movements and activities based on fear, hypervigilance to illness information, muscular reactivity, and disuse/deconditioning, all potentially leading to worse health outcomes (Price 1999). Furthermore, expectations may be associated with changes to treatment received due to modified compliance, overuse, or non-compliance with medications and advice, or changes in health consulting behaviours, which may influence health outcomes.

Alternatively, expectations, which are influenced by what people know about themselves and their circumstances, may reflect at least in part a realistic evaluation of their likely prognosis. This would mean that attempts to modify expectations may constitute false reassurance and, at best, have no impact on outcomes.

Why it is important to do this review

Many primary studies using various research methods, including exploratory and confirmatory study design phases, have investigated the relationship between expectations and low back pain outcomes. The results of several studies suggest an association between expectations and low back pain outcomes. Kapoor 2006 reported that there was a medium to large effect size between negative patient expectations and return-to-work outcomes in an acute low back pain population. Other researchers have observed similar relationships between expectations and return-to-work outcomes in chronic low back pain populations (for example, Hagen 2005; Reme 2009; Sandstrom 1986; Schultz 2005). However, some studies have reported weak or no relationships between expectations and return-to-work outcomes (for example, Gross 2005; Schultz 2002). Gross 2005 found no significant association between work-related recovery expectations and working status at one-year follow-up in a sample with subacute occupational low back pain.

Three focused systematic reviews have synthesised evidence about recovery expectations (Fadyl 2008; Hallegraeff 2012; Iles 2009). Fadyl 2008 reviewed the literature and included 10 studies that investigated how expectations relate to return-to-work outcomes after injuries (including, but not limited to, low back pain). These authors reported that evidence is limited, and they recommended further investigation. Hallegraeff 2012 conducted a review to assess whether negative expectations in people with acute low back pain resulted in increased odds of being off work. Ten studies were included and synthesised; the authors of this review concluded that the odds of not returning to work were twice as high for people with negative recovery expectations. Iles 2009 aimed to determine the predictive strength of negative recovery expectations for the outcome 'activity limitations' in people with acute or subacute non-specific low back pain. The review included 10 studies and reported that recovery expectations measured within the first three weeks of low back pain onset are strong predictors of activity limitations. The literature searches of these reviews are now out of date. Furthermore, existing reviews about

recovery expectations have not explored the impact of different types or measures of expectations, different populations (setting or duration of symptoms or both), or different outcomes (pain, functional limitations, return to work). These factors may explain some inconsistencies of results reported in the literature.

OBJECTIVES

To synthesise evidence on the association between recovery expectations and disability outcomes in adults with low back pain, and explore sources of heterogeneity.

METHODS

We conducted this review within the framework of the Cochrane Back and Neck Group (Furlan 2015) and report it according to PRISMA guidelines (Moher 2009), while supplementing as necessary for a prognostic factor systematic review. Similar to systematic reviews of intervention studies, there are six key steps to prognosis reviews:

1. Defining the review question
2. Identifying studies
3. Selecting studies
4. Critically appraising studies
5. Collecting data
6. Synthesising and interpreting results

We considered each of these steps and used best methods to limit potential biases.

We conducted a focused systematic review (as opposed to a broad review that investigates evidence on many prognostic factors) to facilitate the most complete assessment and interpretation of the evidence available (Hayden 2009).

Criteria for considering studies for this review

Our review includes prognostic study evidence with the definitions of eligible participants (low back pain), the potential prognostic factor of interest (expectations), outcomes, and study designs described below (Table 1).

Types of study designs

We include published reports of prospective and retrospective longitudinal studies investigating the prognosis of low back pain with baseline (defined as each study onset) measurement of participant characteristics and at least three months' follow-up to study participant outcomes. We included publications presenting analyses of randomised controlled trials (RCTs) if they reported on the association between expectations and low back pain outcomes in the study population or a subgroup. We separately describe studies that included treatment effect modification analyses and also met our study selection criteria. We did not include treatment effect modification (interaction) evidence in our syntheses, but included data about the association between expectations and low back pain outcomes when available in these studies.

We separately considered phases of prognostic factor investigation: Phase 1 (exploratory), and Phase 2 (confirmatory) studies, which provide different levels of evidence (Hayden 2008). Exploratory studies identify associations of many potential prognostic factors and outcomes. While exploratory studies are necessary to identify new prognostic factors, they provide the least conclusive information about the independence of a variable as a valid prognostic factor, since results are often presented unadjusted or not adequately adjusted for known covariates. Studies in this exploratory phase of investigation often have widely varying results, as spurious associations are common due to the high number of factors explored, and studies may overstate their conclusions (Hayden 2008). Confirmatory studies, with analyses planned a priori, test the independence of the association between one specific (or just a few) prognostic factor(s) and the outcome of interest. These studies aim to measure the independent (additional) prognostic effect of a factor while controlling for known covariates (i.e. existing or established prognostic factors in the field). We classified included studies according to the authors' objectives and approach to design and analysis, and considered the phase of investigation of studies in our assessment of the strength of the evidence available.

Target population

We included studies involving any population of adult participants with non-specific low back pain, including general populations, occupational, and non-surgical clinical populations. We included studies if they investigated mixed-pain populations (including conditions other than low back pain, such as thoracic or neck pain, or healthy controls) only if the majority (more than 75%) of the study population was experiencing non-specific low back pain, or subgroup information was presented for this population. We included studies where the operationalisation of low back pain was based on symptoms, signs, or consequences of low back pain such as sick leave, medical consultation, or treatment. We included studies with participants at any point in the course of low back pain from acute to subacute/chronic. We aimed to separately consider worker, healthcare and general populations, and explored subgroup analyses with acute (less than six weeks), subacute/chronic (six weeks or more), and mixed-duration low back pain populations. We planned to use sensitivity analyses to explore the robustness of results, excluding studies with mixed pain or specific low back pain populations.

We excluded studies that involved a majority of individuals with low back pain caused by specific pathologies (including nerve root impingement, fracture, ankylosing spondylitis, spondyloarthritis, infection, neoplasm, or metastasis), or specific conditions (for example, pregnancy).

Types of index prognostic factors

We included studies that assessed expectations at baseline or an early point in patient management (i.e. at initial consultation). We defined expectations as 'what participants expect will occur from their low back pain condition'. Included measures of expectations captured two things: 1. individual participant cognition (for example beliefs, perceptions, anticipations, expectations), and 2. related to a future outcome (for example pain, functional limitations, work participation). We separately considered evidence on general expectations, self-efficacy expectations, and treatment expectations, when possible. We excluded current state or trait type of self-efficacy measures, and expectations from outside perspectives (for example, healthcare provider expectations), as well as measures of expected 'process of care' if they did not refer to a future primary outcome of interest.

We included studies investigating treatment expectations if the variable was assessed as a prognostic factor.

We included studies of expectations assessed using any measurement approach: one-dimensional measurement of expectations, for example: “Do you expect that you will have recovered in six months?”, and more complex measurements, for example, using multidimensional validated measurement tools such as the Credibility/Expectancy Questionnaire (Smeets 2008), the Back Pain Self-Efficacy Scale (Levin 1996), or the Pain Self-Efficacy Questionnaire (Nicholas 2007). We used subgroup and sensitivity analyses to explore the impact of different and more robust measurement approaches, and the expectations reference time period (one month, six months, none or unclear reference period), when available.

Types of outcomes

Primary outcomes

We included studies with at least one of the following primary outcomes, according to the International Classification of Functioning, Disability and Health (ICF) framework (WHO 2002):

1. Work participation, measured as return to work, absenteeism, or time on benefits (Steenstra 2012). If multiple measures were available, we selected dichotomous return-to-work measures over time to return to work or time on sick leave;
2. Important recovery in functional limitations, pain intensity (as described below), and/or work participation;
3. Functional limitations, measured by a low back pain-specific scale (for example, the Roland-Morris Disability Questionnaire (RMDQ) (Roland 2000), or the Oswestry Disability Index (ODI) (Fairbank 1980));
4. Pain intensity, measured by a visual analogue scale (VAS) or other pain scale (for example, numerical rating scale (NRS), or McGill pain score (Melzack 1975)).

We recorded study-reported associations of expectations with outcomes analysed using continuous measures of functional limitations or pain intensity (for example, RMDQ on a 24-point scale, or pain VAS on a 10-point scale), and with the measure dichotomised to reflect improvement at the described time points as reported in primary studies ("important recovery"). We included any study-defined definition of improvement, but prioritised and separately considered evidence from studies that used an ideal definition of ‘improvement’ - clinically important individual patient response where improvement in score is 30% or more of its baseline value, with a minimum value of 20-point (/100) in pain and 10-point (/100) in functioning (Kovacs 2007; Ostelo 2008).

We grouped outcome data into four time periods for analyses: short (closest to three months), medium (closest to six months), long follow-up (closest to 12 months), and very long follow-up (more than 16 months). For primary analyses, balancing homogeneity with availability of data, we used available study data from the time period closest to 12 months (defined as ‘long, closest to 12 months’).

Secondary outcomes

We identified the following secondary outcomes, when they were available in included studies:

1. Global improvement or perceived recovery;
2. Health-related quality of life (for example SF-36 (as measured by the general health sub-scale) (Ware 1992), EuroQol (EuroQol Research Foundation 2019), general health (for example, as measured on a VAS scale) or similarly validated index);
3. Satisfaction with treatment;
4. Mood (for example, depression, measured with the Center for Epidemiologic Studies Depression Scale (CES-D) (Radloff 1977));
5. Healthcare use, including costs.

We excluded studies if they did not measure at least one of our primary outcomes; this is justified, as our primary outcomes were selected by our team as the most important patient-oriented low back pain outcomes for prognosis, capturing body function, functional limitation and participation restriction (WHO 2002).

Search methods for identification of studies

The search strategy included electronic searches and additional strategies to retrieve as many relevant publications as possible.

Electronic searches

We conducted focused and broad electronic searches with the help of an experienced Information Scientist, using indexed terms and free-text words, with no date or language restrictions. We searched the following sources from database inception to 12 March 2019.

1. MEDLINE from Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) (1946 to 12 March 2019);
2. Embase from Embase.com (1974 to 12 March 2019);
3. CINAHL from EBSCOhost with Full Text (1981 to 12 March 2019);
4. PsycINFO from EBSCOhost (1887 to 12 March 2019).

Our focused search strategy included terms related to low back pain (Cochrane Back and Neck Group recommended strategy) (Furlan 2015), expectations, and prognostic study methods (prognosis strategy of Wilczynski 2004); see Appendix 1 for the full focused MEDLINE, Embase, CINAHL, and PsycINFO strategies.

We previously observed, in a ‘review of reviews’ on low back pain prognosis, the possible introduction of ‘positive study’ bias in review search strategies that include prognostic factor terms (Hayden 2009). We therefore also included results of a broad search in MEDLINE and Embase (Hayden 2007). This search strategy included terms related to low back pain and prognostic study methods, without focused terms related to expectations (Appendix 2).

Searching other resources

Recognising potential limitations of electronic search strategies, we supplemented our search to identify potentially relevant studies from other sources:

1. Reference searches of relevant reviews, including previously published systematic reviews of expectations and low back pain or musculoskeletal pain (Darlow 2012; Fadyl 2008; Hallegraeff 2012; Iles 2009; Parsons 2007), identified broad systematic reviews of low back pain prognosis or prognostic factors (for example, Haskins 2012; Hendrick 2011; Menezes Costa 2012; Ramond 2011), and reference lists of all included studies for search up to 1 February 2018.

2. Citation searches of relevant recovery expectation measurement tools (Devilly 2000; Levin 1996; Lim 2007; Metcalfe 2005; Nicholas 2007; Sarda 2007; Smeets 2008; Tate 1999).
3. Review of personal files of investigators, which included authors of previous focused reviews of expectations (Iles 2009; Parsons 2007).

The comprehensive search was executed and downloaded into EndNote X8 for electronic bibliographic management.

Data collection

Selection of studies

We used an online electronic systematic review software package (DistillerSR) to organise and track the selection process. Two review authors (from MW, RO, JAH) or other contributor (from MT, JC, AS, EWP) with accelerated screening (i.e. consensus of two review authors required to exclude, one review author required to move a citation forward) conducted initial screening of titles identified through electronic searches, followed by screening titles and abstracts of citations for relevance, using a pre-tested electronic form. We advanced studies if they comprehensively investigated prognostic factors or prediction models associated with any one of our primary outcome measures in a non-specific low back pain population or subgroup. We resolved disagreements by consensus and by recourse to a third review author. We retrieved all articles in full deemed to be relevant, or for which the relevance could not be determined from the abstract.

A challenge in prognostic factor systematic reviews is that determination of whether a study measured a specific prognostic factor often requires full-text screening to avoid potential reporting bias (i.e. a study finding a positive association is more likely to report this association in the title or abstract). We acknowledged this potential bias and screened full-text publications meeting other inclusion criteria using electronic and handsearching. We included low back pain prognostic studies that investigated expectations and their association with at least one of our primary outcomes of interest. Two review authors (from MW, RO, JAH) or other contributor (from MT, JC, AS, EWP) independently confirmed study relevance with the full text, including discussion and consensus with a third review author when necessary.

We linked multiple publications of the same or overlapping participant data as one study, and identified the primary study as the publication presenting the most relevant data for our review question (i.e. reporting the independent association of expectations with low back pain outcome).

Data extraction and management

We extracted data and reached consensus using electronic extraction forms in MS Access and DistillerSR for studies identified in searches to 1 February 2018. We tested and modified the data extraction forms a priori. For each included study we extracted participant characteristics (population source and setting, inclusion criteria, and duration of low back pain episode at baseline), prognostic factor(s) (the expectation constructs as described above, including measurement approach, timing of measurement, prevalence of positive/negative expectations), outcomes (measures assessed and the incidence of poor outcome), study design, follow-up period(s), and all unadjusted and adjusted associations reported between the prognostic factor(s) and outcomes, with details on any adjustment factors used.

If multiple measures of expectations were available in a single study, we extracted information about all measures and associations with outcomes. For primary analyses, we chose the 'best' measurement based on evidence of validity and reliability, and prioritised the order: general, self-efficacy, treatment expectations. If studies presented multiple measures of general expectations with work participation outcomes, we selected the expectations measure referencing a return-to-work event (e.g. return-to-work confidence, expected time to return to work) rather than tied to pain or functional limitations (i.e. risk of persistent pain).

One review authors (MW) and one other contributor (from MT, JC, AS, EWP) independently extracted information; we used a consensus method with a third review author (JAH) consulted in the case of disagreements.

Assessment of risk of bias in included studies

One review author (MW) and one other contributor (from AS, JC, EWP) independently, with consensus and a third review author (JAH) as necessary, assessed each study's risk of bias using the Quality in Prognosis Studies (QUIPS) tool (Hayden 2013), appropriate for prognostic factor review questions. This approach has been recommended by the Cochrane Prognosis Methods Group, used in several reviews (for example, Dawes 2016; Jimenez 2009; Lamberink 2017), and has acceptable inter-rater reliability. We assessed each study's risk of bias considering six domains: study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis and reporting; Appendix 3 presents a copy of the QUIPS tool modified for this review. We describe each of the six domains, paraphrased from Hayden 2013, in Appendix 4. The reviewers considered all available publications of the same study in assessments. We assessed risk of bias for the study overall, based on the primary outcome at highest risk of bias, rather than separately by outcome measure. We were not blinded to study authors, institution, or journal of publication, due to feasibility.

For each of the six domains, we assessed responses to the prompting together (while considering missing or poorly-reported information) to inform the 'Risk of bias' judgement. We recorded information and methodological comments supporting the item assessment, and then judged using the QUIPS tool by rating each domain as having high, moderate, or low risk of bias. We then judged a study's overall validity, and gave it an overall low risk of bias if we had rated all of the six bias domains as having low risk of bias. We used subgroup analyses to explore the impact of biases on the observed size and direction of effect across each of the six 'Risk of bias' domains.

Measures of association to be extracted

We extracted all unadjusted and adjusted measures of association (i.e. prognostic effect estimates) from included studies, and we recorded how expectations were measured and reported. We converted effect sizes, as necessary, to the natural log scale to avoid possible selection bias by allowing us to use data from as many studies as possible. We calculated standard errors (SEs) by log-transforming study-reported confidence intervals (or other measures of variance) and subsequently using an appropriate

conversion formula. We used odds ratios (ORs) in the natural log scale as the common measure of the relationship between expectations and outcome. We used relative risks and hazard ratios (HRs) to estimate ORs (Symons 2002) and we converted standardised regression coefficients for continuous outcomes to natural log ORs for synthesis (Borenstein 2009; Peterson 2005). If available in sufficient numbers, we had planned to separately extract and analyse continuous outcomes on a continuous scale, and HRs for studies providing this measure of association.

When data were available, we separately synthesised adequately-adjusted associations between expectations and low back pain outcomes from unadjusted associations. In our conceptual framework (Figure 1), we defined five domains of other covariates important for this review question: individual demographics, social support, work factors and environment, psychological factors, and low back pain complaint factors. We defined 'minimally adjusted' study analyses as those presenting adjusted analyses controlling for one or two of these domains, and 'adequately adjusted' study analyses as those presenting adjusted analyses controlling for three or more of these domains. For consistency, we recalculated associations to be in the same direction on the natural log scale, as necessary, with effect sizes above 0 indicating that better expectations are associated with a better prognosis.

Studies included in this review collected and analysed the association between expectations and low back pain outcomes at the individual participant level. Some studies presented data stratified for specific characteristics, creating independent subgroups (men and women, or treatment groups). For the three studies where this occurred (Kongsted 2014; Opsahl 2016; Tran 2015), we presented the data separately in meta-analyses, labelled as 'groups'.

Dealing with missing data

We included studies that investigated the relationship between expectations and low back pain outcomes, even if there were incomplete data provided about the size of the effect (for example, if the factor is mentioned only as being 'non-significant' in the analyses, but no information about the size of the effect was reported), or if assumptions were necessary to calculate a measure of variance (for example, if data were incomplete, but the standard error could be calculated from a presented P value). Data reported and necessary data conversions are described in Tables 2 to 5. In sensitivity analyses we excluded studies from meta-analyses if they presented data requiring conversions with uncertain assumptions.

Assessment of heterogeneity

Our secondary objective was to explore sources of heterogeneity to identify the impact of differences in participants, measurement of expectations, low back pain duration, outcome, follow-up length, study design, and risk of bias on the association between expectations and low back pain. We synthesised associations within these clinically-relevant subgroups. To assess statistical heterogeneity across studies included in a particular meta-analysis, we inspected forest plots and quantified heterogeneity using the I^2 statistic and Tau^2 (the estimate of between-study variance).

Assessment of reporting deficiencies

We examined potential publication bias for meta-analyses containing 10 or more studies by visually examining asymmetry in funnel plots, and with Egger's test (Egger 1997). We considered potential publication bias as part of the rating of certainty of the evidence.

Data synthesis

Data synthesis and meta-analysis approaches

We conducted meta-analyses when valid data were available about the prognostic association between expectations and each of our primary outcomes (work participation, important recovery, functional limitations, and pain). We separately synthesised dichotomous and continuous measures of expectations (0 - 10 scale) as they were reported in included studies, as well as for unadjusted and adjusted analyses, when available. To include the most and sufficiently similar studies available, our primary analyses used data from: the longest follow-up period closest to 12 months, the best measure/type of expectations, and the best adjusted model results.

We conducted meta-analyses using Review Manager 5 with a random-effects generic inverse variance meta-analysis model, which accounts for any between-study heterogeneity in the prognostic effect. We pooled effect sizes as natural log ORs and SEs, and converted these pooled estimates to ORs and 95% CIs for ease of interpretation. We present results of forest plots of meta-analyses in the text of this review when three or more studies were available for meta-analyses in primary analyses, and when at least three studies were available for two or more subgroups in subgroup analyses. We considered differences to be statistically significant at the 5% level. We defined the clinical importance of observed associations based on effect size as small (OR < 1.5), moderate (1.5 ≤ OR ≤ 2), or large (OR > 2) (modified from Hartvigsen 2004 and Hemingway 1999).

To allow for fuller interpretation of the evidence available, we also present the results using a narrative approach. For each comparison, we summarise the number of studies that reported positive, neutral or negative associations between expectations and the outcome of interest. Studies reporting a statistically significant relationship between positive expectations and a good outcome were recorded as 'positive'; studies reporting a statistically significant relationship between negative expectations and a good outcome were recorded as 'negative'; we recorded non-significant associations as 'neutral', with moderate or large effect sizes (OR ≥ 1.5) reported as clinically important.

Subgroup analyses and investigation of heterogeneity

For our primary work participation outcome, we explored heterogeneity with subgroup analyses using meta-analyses, and with the narrative synthesis approach. We defined subgroups according to population (acute (less than six weeks), subacute/chronic (six weeks or more) and mixed duration), specific types of measures of expectations (general expectations, self-efficacy expectations, and treatment expectations), the expectations reference time period (short (four weeks or less), long (three to six months), or no/unclear reference period), and outcome measurement (specific follow-up periods). We also conducted separate meta-analyses based on assessments of the study phase (exploratory and confirmatory) and according to risk of bias (by each domain, and overall). We planned to separately consider general, worker, and healthcare source populations.

Sensitivity analyses

We used sensitivity analyses to explore the impact of our judgements of study risk of bias, alternatively including studies rated as low or moderate risk of bias for all domains to indicate overall low risk of bias. Sensitivity analyses also explored the impact of data conversions in cases where we were required to make assumptions about normality and proportionality of data; in sensitivity analysis we excluded studies where the standard error was calculated from a P value. There were not sufficient numbers of studies available to allow other planned sensitivity analyses for studies including only low back pain populations versus studies including a small proportion of mixed pain populations, surgical candidates or individuals with lumbar disc herniation.

Rating of certainty of evidence and 'Summary of findings' tables

We judged and report the overall quality of evidence for our primary outcomes using a modified GRADE (Guyatt 2011) approach that was previously used in another prognostic factor review (Huguet 2013), rating the overall strength of evidence as 'high', 'moderate', 'low' or 'very low', considering phase of investigation, internal validity, size and precision of effect, heterogeneity, generalisability, potential reporting bias, and the size of the observed effect. See Appendix 5, reproduced from Huguet 2013.

RESULTS

Results of the search

We identified 7225 records in our searches: 4807 from electronic database searching, 1469 from citation searches of expectations measures and tools, and 949 from references of other published low back pain prognosis systematic reviews. There were 4635 unique citations, of which we excluded 3727 citations at title/abstract screening. We screened 924 full-text publications (including 16 papers added from reference lists of included studies). We excluded publications after brief full-text screen (798 publications; mainly due to not including a measure of expectations), or with comprehensive review of the full-text publication (29 publications: no appropriate measure of expectations (20), not low back pain (5), primary outcome not available (2), follow-up period not available (2)); see the 'Characteristics of excluded studies' table. We included 60 studies (in 85 publications) that met our inclusion criteria (Figure 2). Sixteen publications are awaiting assessment. In an updated search (12 March 2019) we found 10 studies likely to be eligible for inclusion, and two were publications probably linked to included studies. Four studies were not published in English (Characteristics of studies awaiting classification).

Figure 2. PRISMA flow diagram showing identification and selection of included studies.

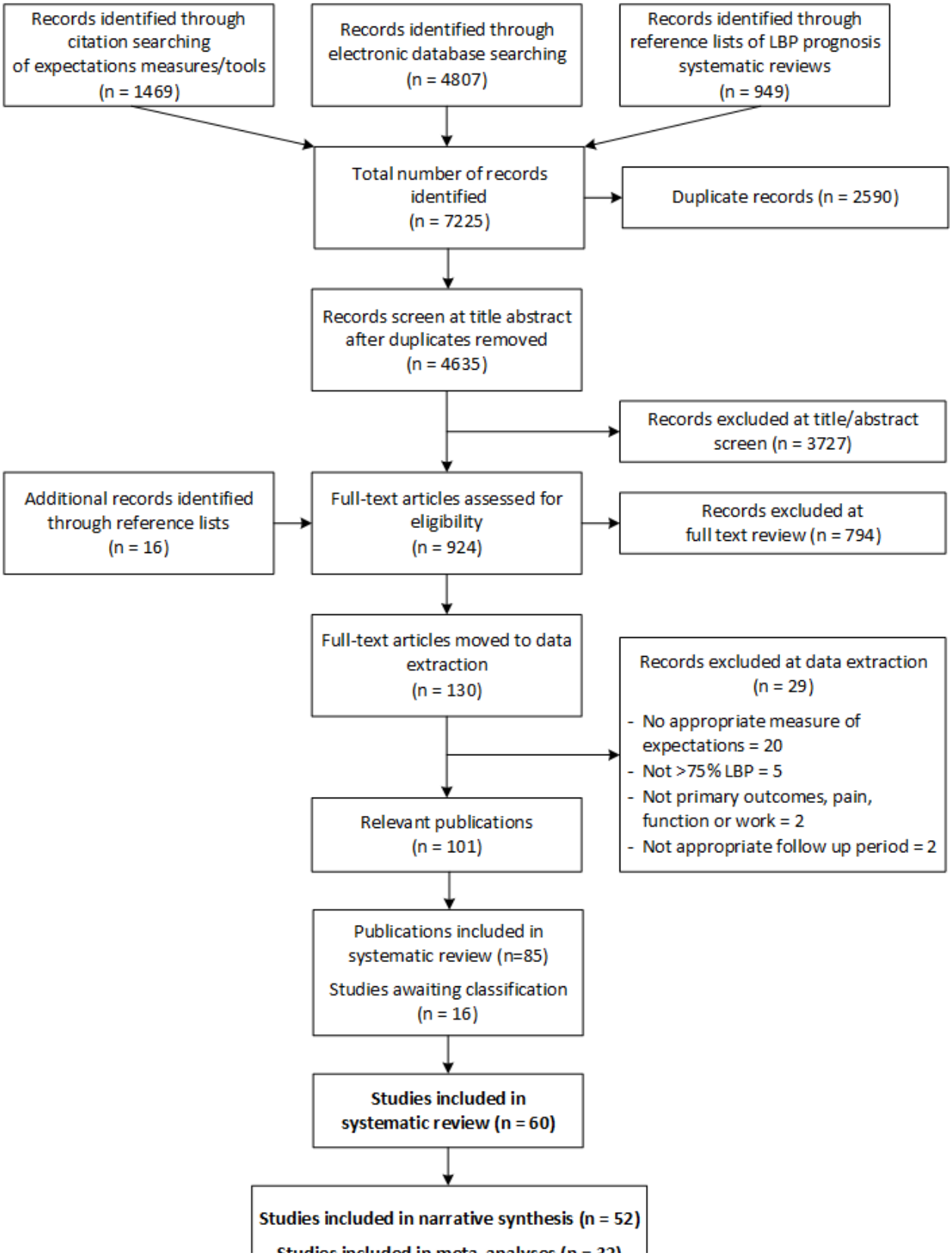
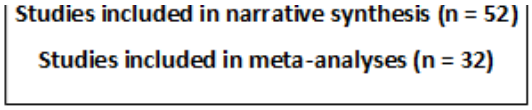


Figure 2. (Continued)



Of the 60 included primary studies reported here, we identified 60% (36 studies) from our electronic searches, 28% (17 studies) from other relevant low back pain prognosis or expectations reviews, 8% (5 studies) from reference searches of included studies, and 3% (2 studies) from searches of expectations measures.

Included studies

Sixty studies (30,530 participants) provided information about expectations and low back pain outcomes and were included in this review (Table 2**:** Descriptive summary of included studies) Beneciuk 2017; Besen 2015; Bishop 2015; Butler 2007; Carriere 2015; Casey 2008; Demmelmaier 2010; Dionne 1997; Downie 2016; Du Bois 2008; Enthoven 2006; Enthoven 2016; Foster 2008; George 2010; Gervais 1991; Glattacker 2013; Goldstein 2002; Gross 2010; Grotle 2006; Haas 2014; Hagen 2005; Haldorsen 1998; Harkapaa 1996; Hazard 1996; Henschke 2008; Heymans 2006; Hildebrandt 1997; Jellema 2002; Jensen 2000; Jensen 2013; Karjalainen 2003; Kongsted 2014; Leboeuf-Yde 2004; Lindell 2010; Macedo 2014; Magnussen 2007; Michaelson 2004; Morlock 2002; Myers 2007; Niemisto 2004; Opsahl 2016; Opsommer 2017; Petersen 2007; Rasmussen-Barr 2012; Reeser 2001; Reiso 2003; Reme 2009; Rundell 2017; Sandstrom 1986; Schultz 2004; Shaw 2009; Sherman 2009; Steenstra 2005; Tran 2015; Truchon 2012; Turner 2008; Underwood 2007; Van Hooff 2014; Van Wijk 2008; Yelland 2006. Forty-four studies (73%) were published more than five years ago (before 2013). Thirty-four studies (56%) were cohort study design and 27 studies (45%) were re-analyses of randomised controlled trials (one study included both, combining data from a cohort study and a randomised controlled trial). The sample size of included studies ranged from 52 (Sandstrom 1986) to 5220 (Rundell 2017), with a median of 257 participants (interquartile range (IQR): 132 to 592). Studies were conducted in the USA (16 studies), Norway (8), Sweden (7), the Netherlands (6), Canada (5), Australia (4), the United Kingdom (4), Finland (3), Denmark (3), Germany (2), Switzerland (1), and Belgium (1).

The study populations had similar numbers of men and women (median, 47% men; IQR 41% to 55%) and mean age ranged from 34 to 74 years. Study populations were mostly chronic (37%; 22 studies) or mixed duration of low back pain (31%; 19 studies), from healthcare (62%; 37 studies) or occupational settings (26%; 16 studies). General expectation was the most common type of expectation measured (70%; 42 studies); 16 studies measured more than one type of expectation. There were 62 distinct measures of expectations used across the 60 studies (46 individual items and 16 multi-item scales). Of general measures, 24 studies referenced expectations of return to work, 21 referenced functioning, recovery, or pain reduction, and three referenced expectations related to multiple study outcomes (i.e. duration of symptoms, activity restrictions and job limitations). Commonly-used measures of expectations included: chance of return to work within six months (9 studies, 7 measured on 0 - 10 scale as Örebro Musculoskeletal Pain Questionnaire (Linton 1998), item 16), perceived risk of pain persistence (9 studies), for example, "In your view, how large is the risk that your current pain may become persistent?" (Örebro Musculoskeletal Pain Questionnaire, item 15 measured on a 0 - 10 scale), and expected helpfulness of treatment (5 studies), for example, "How helpful do you think x treatment will be?" (measured on a 0 - 10 scale in three of the five studies). Our primary outcomes, work participation (58%; 35 studies), functional limitations (60%; 36 studies), and pain intensity (52%; 31 studies), were each assessed in most included studies. Secondary outcomes were reported in a small number of included studies: global improvement (15%; 9 studies), health-related quality of life (10%; 6 studies), satisfaction (8%; 5 studies), mood (5%; 3 studies), healthcare costs (12%; 7 studies). Details of individual study characteristics are described in the Characteristics of included studies.

Forty-seven of the 60 included studies were exploratory phase (44 exploratory, 3 exploratory and treatment effect modification; 78%), and 13 (22%) were confirmatory studies, designed to assess the independent association between expectations and low back pain outcome. Confirmatory-phase evidence was available for all primary outcomes other than important recovery outcomes (for work participation: five confirmatory studies, 1268 participants; functional limitations: four confirmatory studies, 651 participants; pain intensity: five confirmatory studies, 1552 participants).

Of the 60 included studies, 52 had usable data for synthesis (87%; 28,885 participants), reported in 55 separate study groups. The characteristics of the 52 studies were similar to all included studies (Table 2). Reasons that data were not usable for synthesis included: follow-up times more than three months measured but relevant data not presented (4 studies), expectation measures had relevant data reported only as part of a more comprehensive tool (2 studies), expectation measures were used to define clusters for analyses (1 study) and not presenting sufficient data on the association between expectations and low back pain outcome (1 study). There was considerable heterogeneity in measurement of prognostic factors and outcome measures in studies with data available for synthesis. Many studies categorised expectation measures for their main analyses: 21 included studies with usable data reported a dichotomous measure of expectations (40%; 13 studies/14 groups for work participation, six studies for important recovery, three studies for function, one study for pain). We describe measures and prognostic factor cut-points in detail with individual study results for each primary outcome in Table 3; Table 4; Table 5; Table 6.

We analysed and reported the association between expectations and low back pain outcomes as unadjusted results (71%; 37 studies), or adjusted results (85%; 44 studies), or both. Of the 44 studies with adjusted results, we judged 26 to be adequately adjusted (59%), 17 to be minimally adjusted (39%) and one to have unclear adjustment (2%). Two studies (4%) reported only the statistical significance of the association between expectations and low back pain outcome (no measure of effect size for any outcomes), with no mention of clinical significance. There were too few homogeneous studies for synthesis of secondary outcomes.

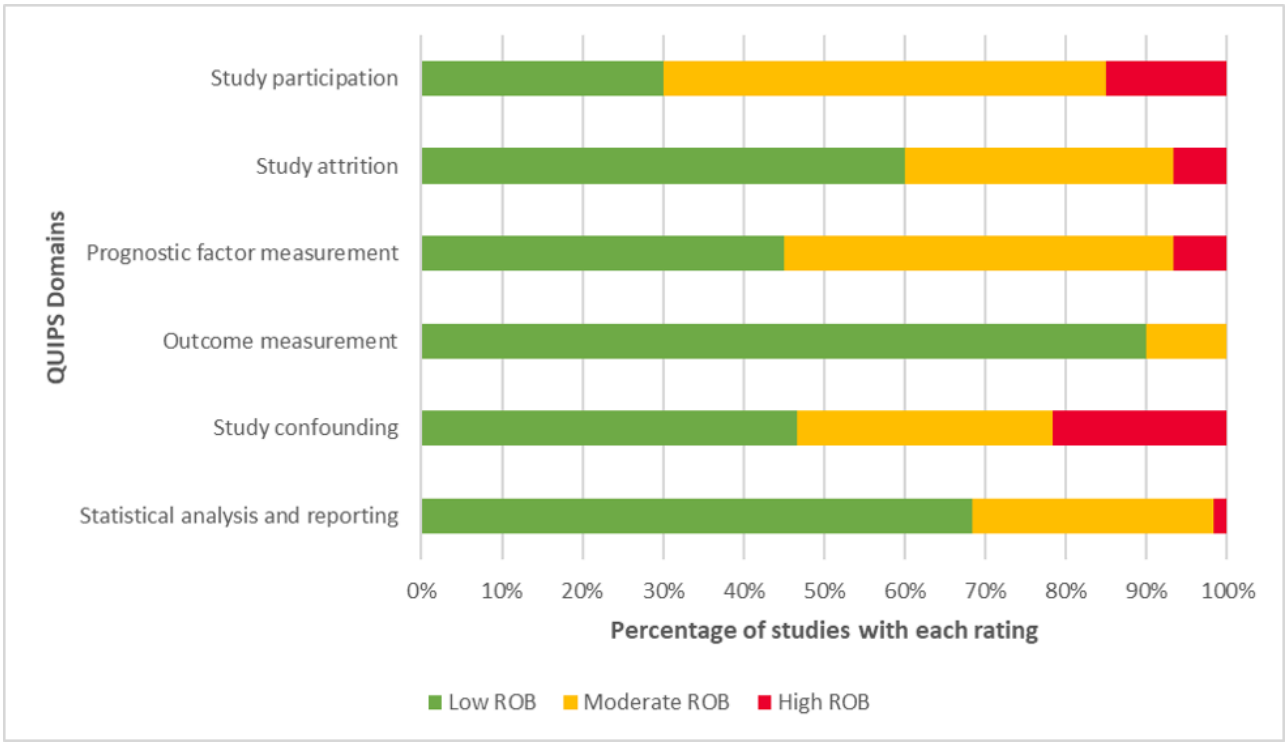
The analytic approach differed considerably across studies, as did the reported effect size types and measures of variance (Table 3; Table 4; Table 5; Table 6). We considered any data presentations where the regression effect and confidence interval and/or standard error and/or exact P value were not fully reported to be incomplete. We considered conversions of odds ratios, risk ratios, hazard ratios and beta coefficients (standardised and non-standardised) to be based on acceptable assumptions, and all other effect conversions to be uncertain. Among all associations reported across outcomes in primary meta-analyses, there were seven removed in a sensitivity analysis from unadjusted data, and one removed from adjusted data.

Risk of bias assessment of included studies

We assessed risks of bias across six domains, using the QUIPS tool for each of the 60 included studies (Table 7). In total, review authors agreed on 278 of 360 items prior to consensus, resulting in a Kappa score of 0.56 (95% CI 0.46 to 0.66). Agreement on the six QUIPS domains ranged from 65% (Prognostic factor measurement) to 88% (Study confounding).

Table 8 presents detailed information about 'Risk of bias' judgements for each included study. We rated 36 studies (60%) overall at low/moderate risk of bias, with all six domains judged to be at low or moderate risk of bias. We assessed only two studies (3%) as having low risk of bias for all of the six domains. Domains where we rated fewer than 50% of studies as having low risk of bias included: Study participation (55% moderate, 15% high risk of bias), Prognostic factor measurement (48% moderate, 7% high risk of bias), and Study confounding (32% moderate, 22% high risk of bias). Figure 3 shows a summary of the review authors' judgements about each 'Risk of bias' domain, presented as percentages across all included studies.

Figure 3. QUIPS risk of bias graph: Review authors' judgements about each risk of bias domain presented as percentages across all included studies (n=60).



Findings

There were between 10 and 25 studies that provided sufficiently similar data about the association between expectations and each of our primary outcomes. There were zero to 12 studies (with 13 groups) available for each of our planned meta-analyses. We report synthesis results in the text only when three or more studies contributed data to a meta-analysis, as described in our protocol. All studies with multiple groups reported associations in the same direction for the multiple groups, so we present narrative syntheses at the study level. Overall, moderate-quality evidence was available for work participation outcome, low-quality evidence for important recovery outcome, very low-quality evidence for functional limitations, and low-quality evidence for pain intensity outcomes (Summary of findings for the main comparison). Among participants with non-specific low back pain, positive expectations were associated with better outcomes for work participation (unadjusted and adjusted); findings for important recovery, functional limitations and pain intensity outcomes are less certain, as described below. No studies reported statistically significant or clinically important negative associations between expectations and any low back pain outcomes. There were insufficient studies presenting HRs to conduct a separate analysis.

Association of expectations with work participation

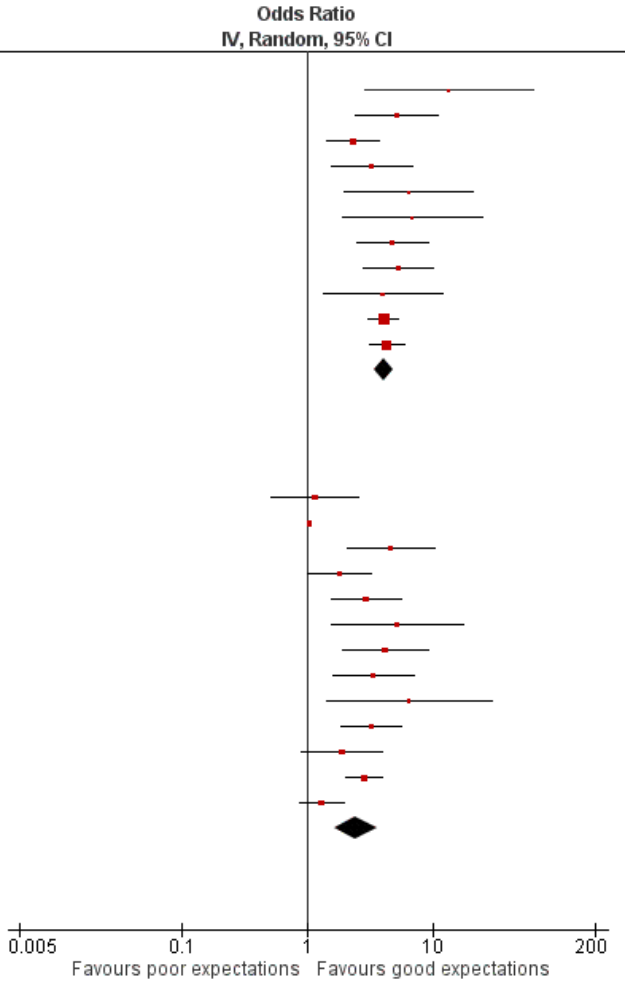
Overall there was moderate-quality evidence for the association of expectations with work participation outcome (Summary of findings for the main comparison). Although there were serious limitations to the data due to moderate risk of bias for many

included studies, and potential publication/reporting bias, the overall quality of evidence was raised by the availability of confirmatory evidence, consistently-reported positive direction of effect, and moderate to large effect size for studies with data available for meta-analysis; there was no serious indirectness of the evidence.

Twenty-five studies (7058 participants) reported information about the association of expectations and work participation (Table 3). Of these, 18 studies (5978 participants) reported unadjusted results, with 17 reporting statistically significant positive associations of expectations with work participation and one study reporting neutral results (i.e. good expectations were statistically significantly associated with better work participation outcomes in almost all available studies). Four studies that reported positive associations for unadjusted analyses did not report adjusted results. Twenty-one studies (6797 participants) reported adjusted results (adjusted for other important covariates), with 16 reporting statistically significant positive associations of expectations with work participation and five studies reporting results that were not statistically significant in either direction (i.e. neutral) (Summary of findings for the main comparison); 11 studies reported clinically important effect sizes.

Fifteen studies (16 groups; 5365 participants) provided sufficiently similar data to allow meta-analysis for work participation outcomes. Pooled unadjusted results found expectations to be associated (statistically significant and clinically important) with work outcomes at the time point closest to 12 months, with expectations measured dichotomously (10 studies, 11 groups; 4528 participants): OR 4.11, 95% CI 3.46 to 4.89; I^2 = 6% (Analysis 1.1; Figure 4). There were similar findings, but with smaller effect sizes (statistically significant and clinically important) and considerable heterogeneity, with pooled adjusted results, with expectations measured dichotomously (12 studies, 13 groups; 4777 participants): OR 2.43, 95% CI 1.64 to 3.62; I^2 = 89%). Only two studies provided sufficient data for unadjusted and adjusted results with expectations measured continuously, not allowing for meaningful meta-analysis (Analysis 1.2). Visual assessment of the funnel plot of adjusted results of the association between dichotomous expectations measures and work participation outcome suggests the presence of small-study effects (Egger's test bias coefficient = 2.79; P < 0.001), which raises the concern of potential publication or reporting bias (Figure 5).

Figure 4. Forest plot: Are expectations associated with work participation (closest to 12 months)? Dichotomous measure of expectations; unadjusted results (10 studies; 11 groups 4,528 participants), and adjusted results (12 studies; 13 groups; 4,777 participants)



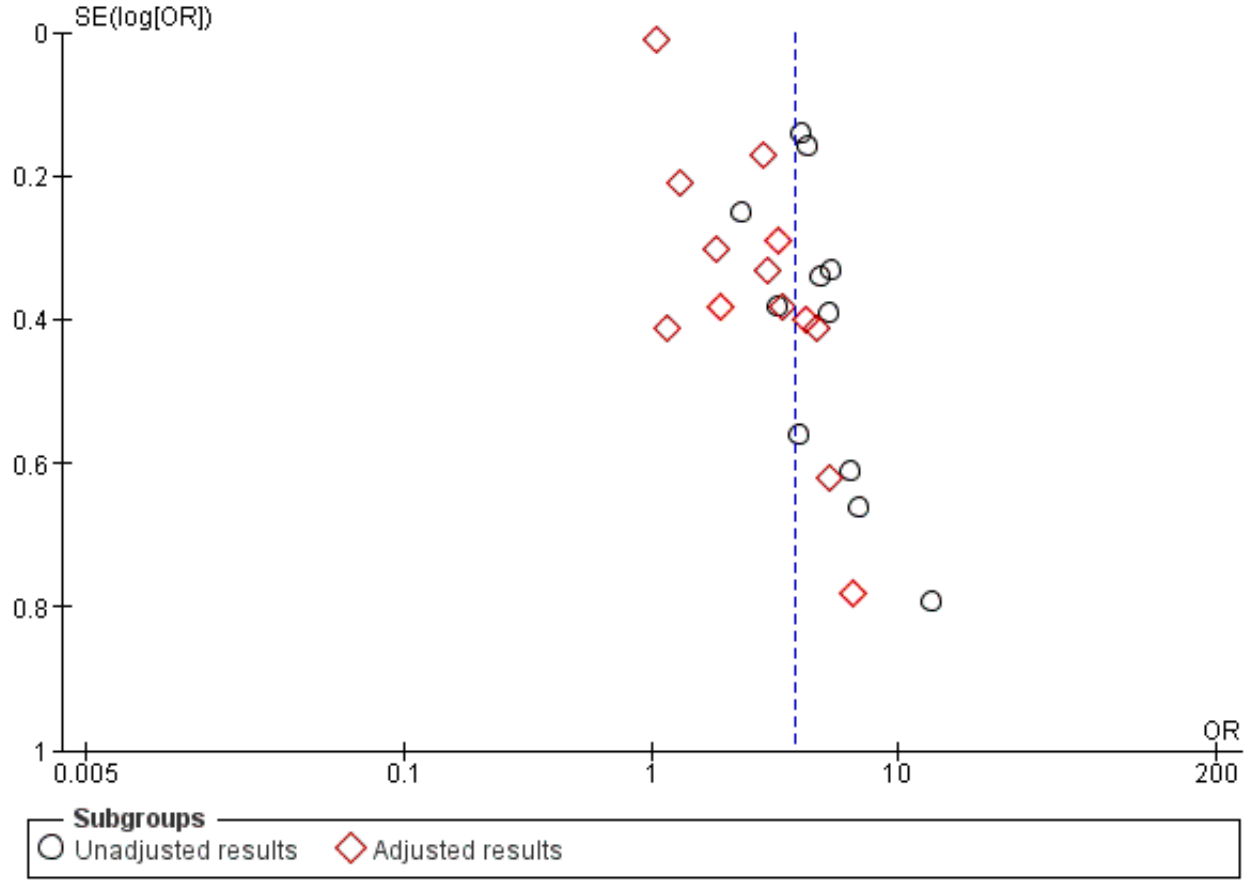
Study or Subgroup	log[Odds Ratio]	SE	Weight	Odds Ratio IV, Random, 95% CI
1.1.1 Unadjusted results				
Carriere 2015 (1)	2.61	0.79	1.2%	13.60 [2.89, 63.97]
Du Bois 2008 (2)	1.65	0.39	5.0%	5.21 [2.42, 11.18]
Hagen 2005 (3)	0.83	0.25	11.6%	2.29 [1.40, 3.74]
Jensen 2013 (4)	1.18	0.38	5.2%	3.25 [1.55, 6.85]
Lindell 2010 (5)	1.86	0.61	2.1%	6.42 [1.94, 21.23]
Magnussen 2007 (6)	1.93	0.66	1.8%	6.89 [1.89, 25.12]
Opsahl 2016 (7)	1.57	0.34	6.5%	4.81 [2.47, 9.36]
Opsahl 2016 (8)	1.68	0.33	6.9%	5.37 [2.81, 10.24]
Reiso 2003 (9)	1.38	0.56	2.5%	3.97 [1.33, 11.91]
Steenstra 2005 (10)	1.4	0.14	31.7%	4.06 [3.08, 5.34]
Turner 2008 (11)	1.46	0.16	25.5%	4.31 [3.15, 5.89]
Subtotal (95% CI)			**100.0%**	**4.11 [3.46, 4.89]**
Heterogeneity: Tau² = 0.01; Chi² = 10.60, df = 10 (P = 0.39); I² = 6%				
Test for overall effect: Z = 15.96 (P < 0.00001)				
1.1.2 Adjusted results				
Butler 2007 (12)	0.14	0.41	7.2%	1.15 [0.52, 2.57]
Carriere 2015 (13)	0.04	0.01	10.2%	1.04 [1.02, 1.06]
Du Bois 2008 (14)	1.53	0.41	7.2%	4.62 [2.07, 10.31]
Hagen 2005 (15)	0.59	0.3	8.4%	1.80 [1.00, 3.25]
Jensen 2013 (16)	1.08	0.33	8.1%	2.94 [1.54, 5.62]
Lindell 2010 (17)	1.65	0.62	5.2%	5.21 [1.54, 17.55]
Opsahl 2016 (18)	1.43	0.4	7.3%	4.18 [1.91, 9.15]
Opsahl 2016 (19)	1.21	0.38	7.5%	3.35 [1.59, 7.06]
Petersen 2007 (20)	1.87	0.78	4.1%	6.49 [1.41, 29.93]
Reiso 2003 (21)	1.17	0.29	8.5%	3.22 [1.83, 5.69]
Reme 2009 (22)	0.64	0.38	7.5%	1.90 [0.90, 3.99]
Steenstra 2005 (23)	1.04	0.17	9.5%	2.83 [2.03, 3.95]
Turner 2008 (24)	0.26	0.21	9.2%	1.30 [0.86, 1.96]
Subtotal (95% CI)			**100.0%**	**2.43 [1.64, 3.62]**
Heterogeneity: Tau² = 0.40; Chi² = 112.62, df = 12 (P < 0.00001); I² = 89%				
Test for overall effect: Z = 4.38 (P < 0.0001)				

Test for subgroup differences: Chi² = 5.60, df = 1 (P = 0.02), I² = 82.1%

Footnotes

(1) 108 participants
(2) 186 participants
(3) 457 participants
(4) 323 participants
(5) 123 participants
(6) 79 participants
(7) 286 participants (women)
(8) 283 participants (men)
(9) 183 participants
(10) 615 participants
(11) 1885 participants
(12) 173 participants
(13) 108 participants
(14) 186 participants
(15) 457 participants
(16) 282 participants
(17) 123 participants
(18) 283 participants (men)
(19) 286 participants (women)
(20) 153 participants
(21) 153 participants
(22) 173 participants
(23) 515 participants
(24) 1885 participants

Figure 5. Funnel plot of comparison: Are expectations associated with work participation (closest to 12 months)? Dichotomous measure of expectations.



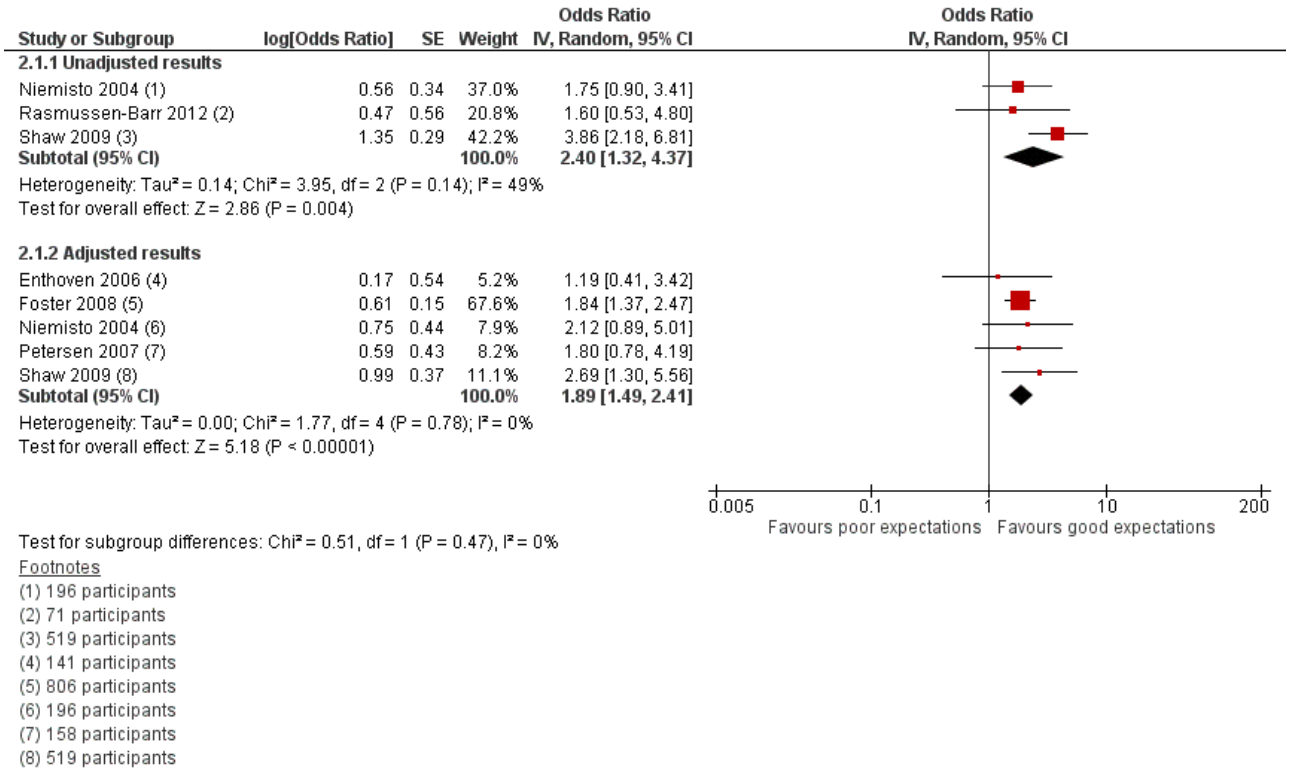
Association of expectations with important recovery

Overall we judged that there was low-quality evidence for the association between expectations and important recovery outcomes. No confirmatory evidence of association was available for this outcome and there were serious limitations to the data, due to moderate risk of bias for many included studies, and potential publication/reporting bias. The overall quality of evidence was raised by a consistently-reported positive direction of effect, and moderate to large effect size for studies with data available for meta-analysis; there was no serious indirectness of the evidence (Summary of findings for the main comparison).

Sixteen studies (10,063 participants) reported usable information about the association of expectations and important recovery (Table 4). Of these studies, 11 (8,872 participants) reported unadjusted results with nine reporting statistically significant positive associations of expectations with important recovery and two studies reporting results that were not statistically significant in either direction. Three studies that reported positive associations for unadjusted analyses did not report adjusted results. Twelve studies (8261 participants) reported adjusted results, with six reporting positive association of expectations with important recovery and six studies reporting results that were not statistically significant in either direction (Summary of findings for the main comparison); six studies reported clinically-important effect sizes.

Ten studies, 10 groups; 7766 participants provided sufficiently similar data allowing meta-analysis for important recovery outcomes. Pooled unadjusted results found expectations to be associated with important recovery (statistically significant and clinically important) at the time point closest to 12 months, with expectations measured dichotomously (3 studies, 3 groups; 786 participants) (OR 2.40, 95% CI 1.32 to 4.37; I^2 = 49%) (Analysis 2.1; Figure 6). We found similar results, but with smaller effect sizes and unimportant heterogeneity, with pooled adjusted results with expectations measured dichotomously (5 studies, 5 groups; 1820 participants) (OR 1.89, 95% CI 1.49 to 2.41; I^2 = 0%) (statistically significant and clinically important), and continuously (4 studies, 4 groups; 1820 participants) (OR 1.15, 95% CI 1.07 to 1.24; I^2 = 81%) (statistically significant, but not clinically important). Only one study provided unadjusted results with expectations measured continuously, not allowing for meaningful meta-analysis (Analysis 2.2).

Figure 6. Forest plot of comparison: 2 Are expectations associated with important recovery (closest to 12 months)?, outcome: 2.1 Dichotomous measure of expectations.



Association of expectations with functional limitations

We judged the overall quality of the evidence to be very low for functional limitations outcome. There were serious limitations to the data, due to moderate risk of bias for many included studies, potential publication/reporting bias, and imprecision of the small estimate for studies available for meta-analysis (Summary of findings for the main comparison).

Thirteen studies (5068 participants) reported usable information about the association of expectations and functional limitations (Table 5). Of these studies, nine (3267 participants) reported unadjusted results, with seven reporting statistically significant positive associations of expectations with functional limitations and two studies reporting results that were not statistically significant in either direction. Two studies that reported positive associations for unadjusted analyses did not report adjusted results. Ten studies (3476 participants) reported adjusted results, with six reporting statistically significant positive association of expectations with functional limitations and four studies reporting results that were not statistically significant in either direction (Summary of findings for the main comparison); seven studies reported clinically important effect sizes.

Seven studies, 8 groups; 3038 participants provided sufficiently similar data to allow meta-analysis for functional limitations outcomes. For dichotomous measures of expectations, only one study provided unadjusted results and two studies provided adjusted results, not allowing for meaningful meta-analysis (Analysis 3.1). Pooled unadjusted results found no relationship between expectations and functional limitations at the time point closest to 12 months with expectations measured continuously (3 studies, 4 groups; 1130 participants) (OR 1.56, 95% CI 0.72 to 3.41; I^2 = 72%). We found similar results with pooled adjusted results with expectations measured continuously (3 studies, 3 groups; 1435 participants) (OR 1.40, 95% CI 0.85 to 2.31; I^2 = 81%) (Analysis 3.2; Summary of findings for the main comparison).

Association of expectations with pain intensity

We judged the overall quality of the evidence to be low for pain intensity outcomes. There were serious limitations to the data due to moderate risk of bias for many included studies, and potential publication/reporting bias. There was a small effect estimate observed from studies available for meta-analysis (Summary of findings for the main comparison).

Ten studies (2900 participants) reported usable information about the association of expectations and pain intensity (Table 6). Of these studies, seven (1853 participants) reported unadjusted results, with four reporting statistically significant positive associations of expectations with improved pain intensity and three studies reporting non-statistically significant, neutral results. Nine studies (2726 participants) reported adjusted results, with five reporting a statistically significant positive association of expectations with improved pain intensity and four studies non-statistically significant, neutral results (Summary of findings for the main comparison); three studies reported clinically important effect sizes.

Four studies (6 groups, 1820 participants) provided sufficiently similar data to allow meta-analysis for pain intensity outcomes.

For dichotomous measures of expectations, only one study (1 group) provided unadjusted results and no study provided adjusted results, not allowing for meaningful meta-analysis (Analysis 4.1). Pooled unadjusted results found expectations not to be significantly associated with pain intensity improvement at the time point closest to 12 months with expectations measured continuously (2 studies, 3 groups, 743 participants) (OR 1.13, 95% CI 0.48 to 2.67; I^2 = 77%). We found evidence of a small prognostic effect with pooled adjusted results with expectations measured continuously (2 studies, 3 groups, 1555 participants) (OR 1.15, 95% CI 1.08 to 1.23; I^2 = 0%) (Analysis 4.2; Summary of findings for the main comparison).

Subgroup and sensitivity analyses

We conducted indirect subgroup comparisons for work participation outcomes at the follow-up period closest to 12 months, using levels of evidence synthesis and meta-analysis for low back pain duration, type of expectation, expectation reference time period, duration of follow-up, exploratory versus confirmatory study design, and for studies judged to be at low or moderate/high risk of bias on each QUIPS domain. For low back pain duration, of the 21 studies (6797 participants) reporting usable adjusted results, one (241 participants) was conducted in an acute low back pain population (positive association of expectations with outcome), nine (2358 participants) in subacute/chronic populations (eight positive, one neutral), and 11 (4198 participants) included mixed-duration low back pain populations (eight positive, four neutral). We found no difference in the effect size for study populations of subacute/chronic low back pain duration (4 studies, 1035 participants) (OR 2.34, 95% CI 1.61 to 3.40; I^2 = 6%) and for study populations with mixed duration of low back pain (8 studies, 9 groups, 3742 participants) (OR 2.41, 95% CI 1.48 to 3.91; I^2 = 91%); test for subgroup effect found no differences (Chi^2 = 0.01; P = 0.93). No studies provided data for meta-analysis in acute low back pain (Analysis 5.1).

No information was available for types of expectation measures other than general recovery. Association of general expectations with work participation outcome was reported with usable adjusted results in 18 studies (6503 participants) (15 with positive association of expectations with outcome, three reporting neutral results), four studies (535 participants) reported associations for self-efficacy expectation measures (two positive, two neutral), and two (375 participants) reported treatment expectations (one positive, one neutral). No studies provided data for meta-analysis of self-efficacy or treatment expectations measures (Analysis 5.2).

For expectations reference time period, of the 21 studies that report useable adjusted results, four (949 participants) referenced one month or less, four (2506 participants) referenced three to six months, and 13 (3342 participants) either referenced a continuum of time or no specific time. We found no significant difference in the adjusted effect sizes for expectations reference periods of one month or less (4 studies, 949 participants) (OR 2.02, 95% CI 1.00 to 4.09; I^2 = 94%), three to six months (4 studies, 2506 participants) (OR 2.83, 95% CI 1.36 to 5.89; I^2 = 75%) and no/unclear reference periods (4 studies, 5 groups, 1322 participants) (OR 2.55, 95% CI 1.53 to 4.25; I^2 = 52%); test for subgroup effect found no differences (Chi^2 = 0.45; P = 0.80) (Analysis 5.3).

For duration of follow-up, five studies (1274 participants) reported short follow-up of approximately three months (four with positive association of expectations with outcome, one with neutral association), six (4438 participants) reported moderate follow-up of approximately six months (three positive, three neutral), 16 (5460 participants) reported long follow-up of approximately 12 months (13 positive, three neutral), and four studies (502 participants) reported very long follow-up greater than 16 months (one positive, three neutral). We found no significant difference in the adjusted effect sizes for follow-up periods: short (3 studies, 816 participants) (OR 3.19, 95% CI 1.77 to 5.75; I^2 = 41%), moderate (4 studies, 2696 participants) (OR 2.43, 95% CI 1.58 to 3.74; I^2 = 45%), long (9 studies, 10 groups, 3923 participants) (OR 2.12, 95% CI 1.41 to 3.17; I^2 = 82%) and very long (2 studies, 276 participants) (OR 3.10, 95% CI 1.87 to 5.12; I^2 = 0%); test for subgroup effect found no differences (Chi^2 = 1.98; P = 0.58) (Analysis 5.4).

For exploratory versus confirmatory phases of study design, of the 21 studies (6797 participants) that reported usable adjusted results, 16 (5529 participants) were conducted using an exploratory approach (11 reported positive association of expectations with outcome, five reported neutral association), and five (1268 participants) were conducted using a confirmatory approach for individual expectations (all six reported positive associations). We found no difference in the effect size between exploratory studies (10 studies, 4100 participants) (OR 2.41, 95% CI 1.76 to 3.29; I^2 = 57%) and confirmatory studies (2 studies, 3 groups, 677 participants) (OR 2.31, 95% CI 0.82 to 6.51; I^2 = 91%); test for subgroup effect found no differences (Chi^2 = 0.01; P = 0.94) (Analysis 5.5).

Results of subgroup analyses for studies rated as low, and moderate/high risk of bias in each QUIPS domain found that the observed direction of association was less consistent (< 75% in the same positive direction) for the Attrition and Confounding domains. The pooled adjusted effect size in studies judged to have low risk of bias due to Attrition was higher (10 studies, 11 groups, 4496 participants) (OR 2.72, 95% CI 2.05 to 3.60; I^2 = 51%) than for studies judged to have moderate/high risk of bias on the same domain (2 studies, 281 participants) (OR 1.04, 95% CI 1.02 to 1.06; I^2 = 0%); test for subgroup effect was statistically significant (Chi^2 = 44.59; P < 0.001), and clinically important (Analysis 6.2). The pooled effect size in studies judged to have low risk of bias due to the Confounding domain was lower (8 studies, 9 groups, 3641 participants) (OR 2.03, 95% CI 1.34 to 3.07; I^2 = 84%) than for studies judged to have moderate/high risk of bias on the same domain (4 studies, 1136 participants) (OR 3.10, 95% CI 2.36 to 4.07; I^2 = 0%) by a clinically important amount, but the test for subgroup differences was not statistically significant (Chi^2 = 2.78; P = 0.10) (Analysis 6.5). Subgroup analyses based on ratings of other QUIPS domains (Participation, Prognostic factor, Outcome, and Analysis and Reporting) found no difference in pooled effect sizes in studies judged to have low risk of bias and studies judged to have moderate/high risk of bias for the same domain (Analysis 6.1; Analysis 6.3; Analysis 6.4; Analysis 6.6).

Removing studies with incomplete data requiring data conversions in five sensitivity analyses (one to two studies removed) made it no longer possible to conduct two meta-analyses of unadjusted results. Removal did not change the conclusions of the meta-

analysis of the association of expectations with work participation outcome.

DISCUSSION

Summary of main results

Our review provides up-to-date evidence about the prognostic association of individual recovery expectations and provides an exemplar for implementation of systematic review methods in prognosis. We included 60 studies, with 52 studies providing data for narrative syntheses, and 24 providing adjusted results for meta-analyses. Despite limitations and heterogeneity of the evidence available, results were quite consistent across primary analyses, as well as subgroup and sensitivity analyses.

We found moderate-quality evidence that positive expectations are strongly associated with better work participation outcomes with large observed association for people with non-specific low back pain at one year. Other outcomes of interest, including important recovery outcomes (moderate observed association); functional limitations and pain intensity outcomes are less certain, respectively achieving low, very low and low quality of evidence. No studies reported statistically significant or clinically important negative associations between recovery expectations and any low back pain outcomes. Our subgroup analyses did not identify any significant differences in the association of recovery expectations and work participation outcome by low back pain duration, recovery expectations reference time period, length of follow-up period or study phase of investigation, although some subgroups had few studies available.

Our results suggest that individual recovery expectations should be considered for inclusion in future prognosis and intervention research studies. There was considerable heterogeneity in how individual studies measured and analysed individual recovery expectations - only eight measures were used in more than one included study. More consistent measurement of patient expectations will facilitate future syntheses. Most studies that we identified used single-item measures of expectations. Single items have been found to be valid and reliable for related constructs (e.g. self-esteem (Robins 2001)), so may be appropriate for measuring expectations as a clear concept with a relevant reference time period and outcome specified. Although not the focus of our review, we observed five multi-item scales of recovery expectations used in included studies that reported evidence of reliability and construct validity (Anderson 1995; Devilly 2000; Nicholas 2007; Shaw 2011; Vlaeyen 1990).

Certainty of evidence available

Overall there was moderate-quality evidence for the association of expectations with work participation. Although there were serious limitations to the data due to moderate risk of bias for many included studies, and potential publication/reporting bias, which is likely to be an important limitation in all prognostic factor systematic reviews, the overall quality of evidence was raised by the availability of confirmatory evidence, consistently-reported positive direction of effect, and moderate to large effect size for studies with data available for meta-analysis; there was no serious indirectness of the evidence. We judged the overall quality of the evidence to be low for important recovery outcomes, very low quality for functional limitations, and low quality for pain intensity outcomes.

Strengths and weaknesses of the review

Our systematic review has numerous strengths. We planned our review a priori with clearly-defined selection criteria. We conducted a comprehensive literature search, using many additional sources to identify relevant studies, including reference searches of other low back pain prognostic factor systematic reviews, expectation measure citation searches, and our broad search of low back pain prognosis studies (limited to the year 2003). Indeed, we identified 38% of included studies from additional search strategies, emphasising the potential importance of such search strategies in prognostic factor reviews. Refining electronic search strategies to improve sensitivity (but not at the expense of specificity) is an area requiring future investigation. We carefully defined characteristics potentially related to heterogeneity, and explored the impact of these with subgroup and sensitivity analyses.

We used both narrative and meta-analytic approaches to synthesise the available evidence, and judged the overall quality of the evidence available. Data for meta-analysis were available from adjusted analyses for only 32 of 60 included studies (53%). Our narrative synthesis approach summarised consistency of results by counting studies that reported positive, neutral or negative associations of expectations with outcome (52 studies, 87%, with available evidence). Although this 'vote counting' approach is often considered a last resort in intervention reviews, we think that it allows for more complete reporting of results when interpreted cautiously alongside meta-analysis. This is particularly important for prognostic factor reviews where reporting bias is likely (studies finding a statistically significant association are more likely to report results) and may lead to an overestimation of effect estimate.

Limitations of our review are mainly related to the quality of the evidence available. We suspect publication or reporting biases, or both, suggesting our results may be overestimated. Positive study bias is likely to be problematic in this (and most other) prognostic factor systematic reviews. Due to feasibility, our literature search for relevant studies included focused searches, i.e. including search terms related to the 'expectations concept' in our electronic search strategy. Studies that report a relationship between the prognostic factor and common outcomes are therefore more likely to have been identified in these searches due to reporting of this positive result in the study abstract. We tried to address this potential bias by using additional search strategies (a broad search of low back pain prognosis studies, citation searches of expectation measures and reference searches of other low back pain prognosis systematic reviews), as well as identifying and including studies that measured expectations, even if it was not reported or included in prognostic analyses (i.e. reported as not statistically significant). We also observed that some studies (fortunately, a small number) reported positive unadjusted association of expectations with an outcome of interest, but then did not report the association adjusted for other important covariates; this further contributes to likely overestimation of our adjusted results. Future work is required to investigate the impact and potential strategies to alleviate publication and reporting bias, as well as initiatives to require protocol registration and publication of prognostic studies (Peat 2014).

Our review found considerable heterogeneity in the measurement and analysis of prognostic factors and outcomes. For our work participation outcome we found increased statistical heterogeneity in the adjusted meta-analysis (I^2 of 89%) compared to the meta-analysis of unadjusted associations (I^2 = 6%); considerable heterogeneity remained in this analysis even when limited to studies that we considered to be ideally adjusted according to our prespecified theoretical framework (I^2 = 84%). This unexpected result emphasises the many sources of heterogeneity in prognostic factor studies. We found expectations measured on a continuous scale to be less strongly associated with outcomes than dichotomous measures of expectations. Different cut-points used to determine 'good' expectations from continuous scales for analyses may be fuelling this strength of association if cut-points were chosen by authors based on statistical significance rather than clinical relevance. There was also heterogeneity in data due to different model-building and covariate adjustment. We have tried to transparently present and explore potential impacts in our review by presenting unadjusted and adjusted results, as is recommended by Riley 2019. We defined a priori an approach to judge studies as being 'minimally' or 'adequately' adjusted and prioritised adequately-adjusted estimates for synthesis. However, overall we judged only 59% of adjusted results to be adequately adjusted, so cautious interpretation of results is warranted. Our 'Risk of bias' and GRADE judgements considered these limitations.

We assessed risks of bias for the study overall, based on the primary outcome at highest risk of bias, rather than separately by outcome measure. This may have overestimated the risks of bias for some outcomes. However, this was relevant for only a small number of studies included in analyses (eight for narrative syntheses, and two for meta-analyses) and did not impact our conclusions. As with any systematic review, when assessing the overall quality of evidence we need to consider many different parts of study design, conduct, and available results. We used an adaptation of GRADE (Huguet 2013) to provide a framework for interpreting overall quality of prognostic factor evidence; this requires further testing and guidance for prognosis reviews.

There were limitations with the internal validity of some included studies for our review purposes. Future prognosis studies in the field should pay particular attention to limit study attrition, improve reporting, collect expectations using more consistent measures, and adequately control for other established prognostic factors. These limitations apply to most reviews of prognostic factors and should be addressed as methods for conducting these types of reviews are developed and refined.

Applicability of findings to clinical practice and policy

The evidence available for this systematic review represents chronic low back pain populations in Europe, North America and Australia, with healthcare and occupational settings equally represented. Only five included studies were conducted in acute low back pain populations, with none reporting data appropriate for our primary meta-analyses of expectations measured dichotomously and return to work. The applicability of our findings for people with acute pain and in developing nations is uncertain.

For applicability of the prognostic factor measures, we aimed to assess evidence about three types of recovery expectations: general expectations, self-efficacy expectations, and treatment expectations. However, most of the evidence available with sufficient homogeneity for meta-analysis was restricted to general expectations. It was not possible to conduct syntheses and subgroup analyses to compare the association of the different types of expectations with low back pain outcomes. The direction of associations for self-efficacy and treatment expectations with outcomes was mostly positive, suggesting similar applicability of findings. Prognostic factor measures used in the included studies were often single items, which are low burden for inclusion in future research studies or subsequent clinical decision-support tools.

Agreements and disagreements with other studies or reviews

Our systematic review provides substantially more evidence on the association between individual recovery expectations and low back pain outcomes than earlier systematic reviews, which were inconclusive, missing relevant studies, and are now out of date. In this systematic review we have drawn similar conclusions to previous reviews (Fadyl 2008; Hallegraeff 2012; Iles 2009), with more confidence supported by a larger number of available studies.

Recent studies investigating the relationship between individual recovery expectations and health outcomes in other fields have reported similar results, including cardiac surgery (Holmes 2016), and major orthopedic trauma outcomes (Busse 2019). A systematic review of measures of patient expectations affecting sick leave reported that people with a range of health conditions (musculoskeletal, mental health, cardiovascular and other) who had lower recovery expectations were less likely to resolve their claim or return to work (Ebrahim 2015).

AUTHORS' CONCLUSIONS

Implications for practice

Expectations are patient-focused, supported by relevant theoretical literature and are potentially modifiable (Iles 2011), providing opportunity for relevant clinical messages and research directions. Our systematic review and meta-analysis found moderate-quality evidence that individual recovery expectations are probably strongly associated (after adjustment) with future work participation outcomes, adding prognostic value. Expectations may be associated with clinically important recovery outcomes (low-quality evidence). The association of recovery expectations with other outcomes of interest is less certain.

Implications for research

Future studies are needed to comprehensively assess psychometric properties of measures of expectations, including comparing reliability and validity of simple single-item measures with multi-item scales, and to examine the impact of including recovery expectations information in clinical management. Prediction models and tools may be further improved for low back pain subgrouping and treatment matching with inclusion. Future studies may include testing the interaction effect between expectations and low back pain outcomes with specific types of treatments. Finally, there is a belief among many back pain stakeholders that expectations are potentially modifiable (Guzman 2007); evidence from our review provides support for further investigation of the effectiveness of new interventions to modify expectations as a mechanism to potentially improve low back pain patient outcomes, particularly for work participation outcomes.

ACKNOWLEDGEMENTS

The authors thank Leah Boulos (Maritime SPOR SUPPORT Unit) and Rachel Couban (formerly with the Cochrane Back Review Group) for their assistance with the search strategies; we thank Andrea Smith, Michelle Tougas, and Jenny Cartwright (formerly with the Department of Community Health & Epidemiology, Dalhousie University) for contributions to early review planning, study screening and data extraction; we thank Emma Wilson-Pease for her assistance with study screening and data extraction; and we appreciate Jenna Ellis's assistance with summary forest plots.

We thank the Cochrane Back and Neck, and Cochrane Prognosis Methods Group peer reviewers and editors for their comments and helpful suggestions, including: Chris Maher, Professor, University of Sydney and Director, Institute for Musculoskeletal Health, Sydney; Tim Carey MD MPH University of North Carolina at Chapel Hill; Dr Marialena Trivella, Centre for Statistics in Medicine, University of Oxford; and Janet Gunderson.

REFERENCES

References to studies included in this review

Beneciuk 2017 *{published data only}*

Beneciuk JM, Hill JC, Campbell P, Afolabi E, George SZ, Dunn KM, et al. Identifying treatment effect modifiers in the STarT Back trial: A secondary analysis. *Journal of Pain* 2017;**18**(1):54-65.

Besen 2015 *{published data only}*

* Besen E, Young AE, Shaw WS. Returning to work following low back pain: towards a model of individual psychosocial factors. *Journal of Occupational Rehabilitation* 2015;**25**(1):25-37.

Carstens JK, Shaw WS, Boersma K, Reme SE, Pransky G, Linton SJ. When the wind goes out of the sail - declining recovery expectations in the first weeks of back pain. *European Journal of Pain (London, England)* 2014;**18**(2):269-78.

Reme SE, Shaw WS, Steenstra IA, Woiszwillo MJ, Pransky G, Linton SJ. Distressed, immobilized, or lacking employer support? A sub-classification of acute work-related low back pain. *Journal of Occupational Rehabilitation* 2012;**22**(4):541-52.

Bishop 2015 *{published data only}*

Bishop FL, Yardley l, Prescott P, Cooper C, Little P, Lewith GT. Psychological covariates of longitudinal changes in back-related disability in patients undergoing acupuncture. *Clinical Journal of Pain* 2015;**31**(3):254-64.

Butler 2007 *{published data only}*

Baldwin ML, Butler RJ, Johnson WG. Self-reported severity measures as predictors of return-to-work outcomes in occupational back pain. *Journal of Occupational Rehabilitation* 2007;**17**(4):683-700.

* Butler RJ, Johnson WG. It pays to be nice: employer-worker relationships and the management of back pain claims. *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine* 2007;**49**(2):214-25.

Butler RJ, Johnson WG, Gray BP. Timing makes a difference: early nurse case management intervention and low back pain. *Professional Case Management* 2007;**12**(6):316-27.

Carriere 2015 *{published data only}*

Carriere J, Thibault P, Sullivan M. The mediating role of recovery expectancies on the relation between depression and return-to-work. *Journal of Occupational Rehabilitation* 2015;**25**(2):348-56.

Casey 2008 *{published data only}*

Casey CY, Greenberg MA, Nicassio PM, Harpin RE, Hubbard D. Transition from acute to chronic pain and disability: A model including cognitive, affective, and trauma factors.. *Pain* 2008;**134**(1-2):69-79.

Demmelmaier 2010 *{published data only}*

Demmelmaier I, Asenlof P, Lindberg P, Denison E. Biopsychosocial predictors of pain, disability, health care consumption, and sick leave in first-episode and long-term back pain: a longitudinal study in the general population. *International Journal of Behavioral Medicine* 2010;**17**(2):79-89.

Dionne 1997 *{published data only}*

Dionne CE, Koepsell TD, Von Korff M, Deyo RA, Barlow WE, Checkoway H. Predicting long-term functional limitations among back pain patients in primary care settings. *Journal of Clinical Epidemiology* 1997;**50**(1):31-43.

Downie 2016 *{published data only}*

Downie AS, Hancock MJ, Rzewuska M, Williams CM, Lin CW, Maher CG. Trajectories of acute low back pain: a latent class growth analysis. *Pain* 2016;**157**(1):225-34.

Du Bois 2008 *{published data only}*

Du Bois M, Donceel P. A screening questionnaire to predict no return to work within 3 months for low back pain claimants. *European Spine Journal* 2008;**17**(3):380-5. [PUBMED: 18172698]

Enthoven 2006 *{published data only}*

Enthoven P, Skargren E, Carstensen J, Oberg B. Predictive factors for 1-year and 5-year outcome for disability in a working population of patients with low back pain treated in primary care. *Pain* 2006;**122**(1-2):137-44.

Enthoven 2016 *{published data only}*

* Enthoven WT, Koes BW, Bierma-Zeinstra SM, Bueving HJ, Bohnen AM, Peul WC, et al. Defining trajectories in older adults with back pain presenting in general practice. *Age and Ageing* 2016;**45**(6):878-83.

Scheele J, Enthoven WT, Bierma-Zeinstra SM, Peul WC, Van Tulder MW, Bohnen AM, et al. Course and prognosis of older back pain patients in general practice: A prospective cohort study. *Pain* 2013;**154**(6):951-7.

Foster 2008 *{published data only}*

Campbell P, Foster NE, Thomas E, Dunn KM. Prognostic indicators of low back pain in primary care: five-year prospective study. *Journal of Pain* 2013;**14**(8):873-83.

* Foster NE, Bishop A, Thomas E, Main C, Horne R, Weinman J, et al. Illness perceptions of low back pain patients in primary care: what are they, do they change and are they associated with outcome?. *Pain* 2008;**136**(1-2):177-87.

Foster NE, Thomas E, Bishop A, Dunn KM, Main CJ. Distinctiveness of psychological obstacles to recovery in low back pain patients in primary care. *Pain* 2010;**148**(3):398-406.

Vergouw D, Heymans MW, Foster NE, Dunn KM, Horst HE. Missing data and imputation: A practical illustration in a prognostic study on low back pain. *Journal of Manipulative and Physiological Therapeutics* 2012;**35**(6):464-71.

George 2010 *{published data only}*

George SZ, Robinson ME. Preference, expectation, and satisfaction in a clinical trial of behavioral interventions for acute and sub-acute low back pain. *Journal of Pain* 2010;**11**(11):1074-82.

Gervais 1991 *{published data only}*

Gervais S, Dupuis G, Veronneau F, Bergeron Y, Millette D, Avard J. Predictive model to determine cost/benefit of early detection and intervention in occupational low back pain. *Journal of Occupational Rehabilitation* 1991;**1**(2):113-31.

Glattacker 2013 *{published data only}*

Glattacker M, Heyduck K, Meffert C. Illness beliefs and treatment beliefs as predictors of short-term and medium-term outcome in chronic back pain. *Journal of Rehabilitation Medicine* 2013;**45**(3):268-76.

Goldstein 2002 *{published data only}*

Goldstein MS, Morgenstern H, Hurwitz EL, Yu F. The impact of treatment confidence on pain and related disability among patients with low-back pain: results from the University of California, Los Angeles, low-back pain study. *Spine Journal* 2002;**2**(6):391-9.

Gross 2010 *{published data only}*

Gross DP, Battie MC. Functional capacity evaluation performance does not predict sustained return to work in claimants with chronic back pain. *Journal of Occupational Rehabilitation* 2005;**15**(3):285-94.

* Gross DP, Battie MC. Recovery expectations predict recovery in workers with back pain but not other musculoskeletal conditions. *Journal of Spinal Disorders and Techniques* 2010;**23**(7):451-6.

Gross DP, Battie MC. Work-related recovery expectations and the prognosis of chronic low back pain within a workers' compensation setting. *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine* 2005;**47**(4):428-33.

Grotle 2006 *{published data only}*

Grotle M, Vollestad NK, Brox JI. Screening for yellow flags in first-time acute low back pain: reliability and validity of a Norwegian version of the Acute Low Back Pain Screening Questionnaire. *Clinical Journal of Pain* 2006;**22**(5):458-67.

Haas 2014 *{published data only}*

Haas M, Vavrek D, Neradilek MB, Polissar N. A path analysis of the effects of the doctor-patient encounter and expectancy in an open-label randomized trial of spinal manipulation for the care of low back pain. *BMC Complementary and Alternative Medicine* 2014;**14**:16.

Hagen 2005 *{published data only}*

Hagen EM, Svensen E, Ericksen HR. Predictors and modifiers of treatment effect influencing sick leave in subacute low back pain patients. *Spine* 2005;**30**(24):2717-23.

Haldorsen 1998 *{published data only}*

Haldorsen EM, Indahl A, Ursin H. Patients with low back pain not returning to work - A 12-month follow-up study. *Spine* 1998;**23**(11):1202-7.

* Haldorsen EM, Kronholm K, Skouen JS, Ursin H. Predictors for outcome of a multi-modal cognitive behavioural treatment program for low back pain patients - a 12-month follow-up study. *European Journal of Pain* 1998;**2**(4):293-307.

Harkapaa 1996 *{published data only}*

Harkapaa K, Jarvikoski A, Estlander A. Health optimism and control beliefs as predictors for treatment outcome of a multimodal back treatment program. *Psychology & Health* 1996;**12**(1):123-34.

Hazard 1996 *{published data only}*

Hazard RG, Haugh LD, Reid S, Preble JB, MacDonald L. Early prediction of chronic disability after occupational low back injury. *Spine* 1996;**21**(8):945-51.

Henschke 2008 *{published data only}*

Costa L, Maher CG, McAuley JH, Hancock MJ, Herbert RD, Refshauge KM, et al. Prognosis for patients with chronic low back pain: inception cohort study. *BMJ* 2009;**339**:b3829.

* Henschke N, Maher CG, Refshauge KM, Herbert RD, Cumming RG, Bleasel J, et al. Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *BMJ* 2008;**337**(7662):a171.

Heymans 2006 *{published data only}*

Heymans MW, Anema JR, Van Buuren S, Knol DL, Van Mechelen W, De Vet HC. Return to work in a cohort of low back pain patients: development and validation of a clinical prediction rule. *Journal of Occupational Rehabilitation* 2009;**19**(2):155-65.

* Heymans MW, De Vet HC, Knol DL, Bongers PM, Koes BW, Van Mechelen W. Workers' beliefs and expectations affect return to work over 12 months. *Journal of Occupational Rehabilitation* 2006;**16**(4):685-95.

Hildebrandt 1997 *{published data only}*

Hildebrandt J, Pfingsten M, Saur P, Jansen J. Prediction of success from a multidisciplinary treatment program for chronic low back pain. *Spine* 1997;**22**(9):990-1001.

Jellema 2002 *{published data only}*

Jellema P, Bierma-Zeinstra SM, Van Poppel MN, Bernsen RM, Koes BW. Feasibility of lumbar supports for home care workers with low back pain. *Occupational Medicine (Oxford)* 2002;**52**(6):317-23.

Jensen 2000 *{published data only}*

Jensen IB, Bodin L, Ljungqvist T, Gunnar Bergström K, Nygren A. Assessing the needs of patients in pain: a matter of opinion?. *Spine* 2000;**25**(21):2816-23.

Jensen 2013 *{published data only}*

Jensen OK, Stengaard-Pedersen K, Jensen C, Nielsen CV. Prediction model for unsuccessful return to work after hospital-based intervention in low back pain patients. *BMC Musculoskeletal Disorders* 2013;**14**:140.

Karjalainen 2003 *{published data only}*

Karjalainen K, Malmivaara A, Mutanen P, Pohjolainen T, Roine R, Hurri H. Outcome determinants of subacute low back pain. *Spine* 2003;**28**(23):2634-40.

Kongsted 2014 *{published data only}*

* Kongsted A, Vach W, Axo M, Bech RN, Hestbaek L. Expectation of recovery from low back pain: a longitudinal cohort study investigating patient characteristics related to expectations and the association between expectations and 3-month outcome. *Spine* 2014;**39**(1):81-90.

Molgaard Nielsen A, Hestbaek L, Vach W, Kent P, Kongsted A. Latent class analysis derived subgroups of low back pain patients - do they have prognostic capacity?. *BMC Musculoskeletal Disorders* 2017;**18**(1):345.

Leboeuf-Yde 2004 *{published data only}*

Leboeuf-Yde C, Gronstvedt A, Borge JA, Lothe J, Magnesen E, Nilsson O, et al. The Nordic Back Pain Subpopulation Program: Demographic and clinical predictors for outcome in patients receiving chiropractic treatment for persistent low-back pain. *Journal of Manipulative and Physiological Therapeutics* 2004;**27**(8):493-502.

Lindell 2010 *{published data only}*

Lindell O, Johansson SE, Strender LE. Predictors of stable return-to-work in non-acute, non-specific spinal pain: low total prior sick-listing, high self prediction and young age. A two-year prospective cohort study. *BMC Family Practice* 2010;**11**:53.

Macedo 2014 *{published data only}*

* Macedo LG, Maher CG, Hancock MJ, Kamper SJ, McAuley JH, Stanton TR, et al. Predicting response to motor control exercises and graded activity for patients with low back pain: preplanned secondary analysis of a randomized controlled trial. *Physical Therapy* 2014;**94**(11):1543-54.

Macedo LG, Maher CG, Latimer J, McAuley JH, Hodges PW, Rogers WT. Nature and determinants of the course of chronic low back pain over a 12-month period: A cluster analysis. *Physical Therapy* 2014;**94**(2):210-21.

Magnussen 2007 *{published data only}*

Magnussen L, Strand LI, Skouen JS, Eriksen HR. Motivating disability pensioners with back pain to return to work -- a randomized controlled trial. *Journal of Rehabilitation Medicine (Stiftelsen Rehabiliteringsinformation)* 2007;**39**(1):81-7.

Michaelson 2004 *{published data only}*

Michaelson P, Sjolander P, Johansson H. Factors predicting pain reduction in chronic back and neck pain after multimodal treatment. *Clinical Journal of Pain* 2004;**20**(6):447-54.

Morlock 2002 *{published data only}*

Morlock RJ, Lafata JE, Nerenz D, Schiller M, Rosenblum M. Expectations, outcomes, and medical costs in patients with low back pain referred to physical therapy. *Disease Management* 2002;**5**(4):185-8.

Myers 2007 *{published data only}*

Myers SS, Phillips RS, Davis RB, Cherkin DC, Legedza A, Kaptchuk TJ, et al. Patient expectations as predictors of outcome in patients with acute low back pain. *Journal of General Internal Medicine* 2008;**23**(2):148-53.

Niemisto 2004 *{published data only}*

Niemisto L, Sarna S, Lahtinen-Suopanki T, Lindgren KA, Hurri H. Predictive factors for 1-year outcome of chronic low back pain following manipulation, stabilizing exercises, and physician consultation or physician consultation alone. *Journal of Rehabilitation Medicine* 2004;**36**(3):104-9.

Opsahl 2016 *{published data only}*

Opsahl J, Eriksen HR, Tveito TH. Do expectancies of return to work and Job satisfaction predict actual return to work in workers with long lasting LBP?. *BMC Musculoskeletal Disorders* 2016;**17**(1):1-10.

Opsommer 2017 *{published data only}*

Opsommer E, Rivier G, Crombez G, Hilfiker R. The predictive value of subsets of the Orebro Musculoskeletal Pain Screening Questionnaire for return to work in chronic low back pain. *European Journal of Physical and Rehabilitation Medicine* 2017;**53**(3):359-65.

Petersen 2007 *{published data only}*

Petersen T, Larsen K, Jacobsen S. One-year follow-up comparison of the effectiveness of McKenzie treatment and strengthening training for patients with chronic low back pain: outcome and prognostic factors. *Spine* 2007;**32**(26):2948-56.

Rasmussen-Barr 2012 *{published data only}*

Rasmussen-Barr E, Campello M, Arvidsson I, Nilsson-Wikmar L. Factors predicting clinical outcome 12 and 36 months after an exercise intervention for recurrent low-back pain. *Disability and Rehabilitation* 2012;**34**(2):136-44.

Reeser 2001 *{published data only}*

Reeser JC, Wiegmann SM, Hoover N, Oldridge N, Phillips D, Bjelland T, et al. Treatment of acute low back pain in Wisconsin: results of the State Medical Society. *WMJ : official publication of the State Medical Society of Wisconsin* 2001;**100**(8):35-42.

Reiso 2003 *{published data only}*

Reiso H, Nygard JF, Jorgensen GS, Holanger R, Soldal D, Bruusgaard D. Back to work: predictors of return to work among patients with back disorders certified as sick: a two-year follow-up study. *Spine* 2003;**28**(13):1468-73; discussion 1473-4. [PUBMED: 12838108]

Reme 2009 *{published data only}*

Reme SE, Hagen EM, Eriksen HR. Expectations, perceptions, and physiotherapy predict prolonged sick leave in subacute low back pain. *BMC Musculoskeletal Disorders* 2009;**10**:139.

Rundell 2017 *{published data only}*

Rundell S, Sherman K, Heagerty P, Mock C, Jarvik J. The clinical course of pain and function in older adults with a new primary care visit for back pain. *Journal of the American Geriatrics Society* 2015;**63**(3):524-30.

* Rundell SD, Sherman KJ, Heagerty PJ, Mock CN, Dettori NJ, Comstock BA, et al. Predictors of persistent disability and back pain in older adults with a new episode of care for back pain. *Pain Medicine* 2017;**18**(6):1049-62.

Sandstrom 1986 *{published data only}*

Sandstrom J, Esbjornsson E. Return to work after rehabilitation. The significance of the patient's own prediction. *Scandinavian Journal of Rehabilitation Medicine* 1986;**18**(1):29-33.

Schultz 2004 *{published data only}*

Schultz IZ, Crook J, Berkowitz J, Milner R, Meloche GR. Predicting return to work after low back injury using the psychosocial risk for occupational disability instrument: A validation study. *Journal of Occupational Rehabilitation* 2005;**15**(3):365-76.

* Schultz IZ, Crook J, Meloche GR, Berkowitz J, Milner R, Zuberbier OA, et al. Psychosocial factors predictive of occupational low back disability: towards development of a return-to-work model. *Pain* 2004;**107**(1-2):77-85.

Schultz IZ, Crook JM, Berkowitz J, Meloche GR, Milner R, Zuberbier OA, et al. Biopsychosocial multivariate predictive model of occupational low back disability. *Spine* 2002;**27**(23):2720-5.

Shaw 2009 *{published data only}*

Kapoor S, Shaw WS, Pransky G, Patterson W. Initial patient and clinician expectations of return to work after acute onset of work-related low back pain. *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine* 2006;**48**(11):1173-80.

Shaw WS, Pransky G, Patterson W, Linton SJ, Winters T. Patient clusters in acute, work-related back pain based on patterns of disability risk factors. *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine* 2007;**49**(2):185-93.

* Shaw WS, Pransky G, Winters T. The Back Disability Risk Questionnaire for work-related, acute back pain: prediction of unresolved problems at 3-month follow-up. *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine* 2009;**51**(2):185-94.

Shaw WS, Reme SE, Linton SJ, Huang YH, Pransky G. 3rd place, PREMUS best paper competition: development of the return-to-work self-efficacy (RTWSE-19) questionnaire--psychometric properties and predictive validity. *Scandinavian Journal of Work, Environment & Health* 2011;**37**(2):109-19.

Sherman 2009 *{published data only}*

* Sherman KJ, Cherkin DC, Ichikawa L, Avins AL, Barlow WE, Khalsa PS, et al. Characteristics of patients with chronic back pain who benefit from acupuncture. *BMC Musculoskeletal Disorders* 2009;**10**:114.

Sherman KJ, Cherkin DC, Ichikawa L, Avins AL, Delaney K, Barlow WE, et al. Treatment expectations and preferences as predictors of outcome of acupuncture for chronic back pain. *Spine* 2010;**35**(15):1471-7.

Steenstra 2005 *{published data only}*

Steenstra IA, Koopman FS, Knol DL, Kat E, Bongers PM, De Vet HC, et al. Prognostic factors for duration of sick leave due to low-back pain in Dutch health care professionals. *Journal of Occupational Rehabilitation* 2005;**15**(4):591-605.

Tran 2015 *{published data only}*

Tran HH, Weinberg J, Sherman KJ, Saper RB. Preference and expectation for treatment assignment in a randomized controlled trial of once- vs twice-weekly yoga for chronic low back pain. *Global Advances in Health and Medicine* 2015;**4**(1):34-9.

Truchon 2012 *{published data only}*

Truchon M, Schmouth M, Fillion AL, Rossignol M, Durand M. Absenteeism screening questionnaire (ASQ): a new tool for predicting long-term absenteeism among workers with low back pain. *Journal of Occupational Rehabilitation* 2012;**22**(1):27-50.

Turner 2008 *{published data only}*

Franklin GM, Rahman EA, Turner JA, Daniell WE, Fulton-Kehoe D. Opioid use for chronic low back pain: A prospective, population-based study among injured workers in Washington State, 2002-2005. *Clinical Journal of Pain* 2009;**25**(9):743-51.

Fulton-Kehoe D, Stover BD, Turner JA, Sheppard L, Gluck JV, Wickizer TM, et al. Development of a brief questionnaire to predict long-term disability. *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine* 2008;**50**(9):1042-52.

* Turner JA, Franklin G, Fulton-Kehoe D, Sheppard L, Stover B, Wu R, et al. ISSLS Prize Winner: Early predictors of chronic work disability: a prospective, population-based study of workers with back injuries. *Spine* 2008;**33**(25):2809-18.

Turner JA, Franklin G, Fulton-Kehoe D, Sheppard L, Wickizer TM, Wu R, et al. Worker recovery expectations and fear-avoidance predict work disability in a population-based workers' compensation back pain sample. *Spine* 2006;**31**(6):682-9.

Underwood 2007 *{published data only}*

* Underwood MR, Morton, Farrin A, UK BEAM Trial Team. Do baseline characteristics predict response to treatment for low back pain? Secondary analysis of the UK BEAM dataset.. *Rheumatology* 2007;**46**(8):1297-302.

Van Hooff 2014 *{published data only}*

Van Hooff ML, Spruit M, O'Dowd JK, Van Lankveld W, Fairbank JC, Van Limbeek J. Predictive factors for successful clinical outcome 1 year after an intensive combined physical and psychological programme for chronic low back pain. *European Spine Journal* 2014;**23**(1):102-12.

Van Wijk 2008 *{published data only}*

Van Wijk RM, Geurts JW, Lousberg R, Wynne HJ, Hammink E, Knape Johannes TA, et al. Psychological predictors of substantial pain reduction after minimally invasive radiofrequency and injection treatments for chronic low back pain. *Pain Medicine (Malden, Mass.)* 2008;**9**(2):212-21.

Yelland 2006 *{published data only}*

Yelland MJ, Schluter PJ. Defining worthwhile and desired responses to treatment of chronic low back pain. *Pain Medicine (Malden, Mass.)* 2006;**7**(1):38-45.

References to studies excluded from this review

Alyousef 2018 *{published data only}*

Alyousef B, Cicuttini F, Davis S, Bell R, Botlero R, Urquhart D. Negative beliefs about back pain are associated with persistent, high levels of low back pain and disability in community-based women. *Menopause* 2018;**25**:977-84.

De Schepper 2016 *{published data only}*

De Schepper EI, Koes BW, Oei EH, Bierma-Zeinstra SM, Luijsterburg PA. The added prognostic value of MRI findings for recovery in patients with low back pain in primary care: a 1-year follow-up cohort study. *European Spine Journal* 2016;**25**(4):1234-41.

Dozois 1995 *{published data only}*

Dozois DJ, Dobson KS, Wong M, Hughes D, Long A. Factors associated with rehabilitation outcome in patients with low back pain (LBP): Prediction of employment outcome at 9-month follow-up. *Rehabilitation Psychology* 1995;**40**(4):243-59.

Du Bois 2009 *{published data only}*

Du Bois M, Szpalski M, Donceel P. Patients at risk for long-term sick leave because of low back pain. *Spine Journal* 2009;**9**(5):350-9.

Elfering 2015 *{published data only}*

Elfering A, Muller U, Rolli Salathe C, Tamcan O, Mannion AF. Pessimistic back beliefs and lack of exercise: a longitudinal risk study in relation to shoulder, neck, and back pain. *Psychology, Health & Medicine* 2015;**20**(7):767-80.

Estlander 1998 *{published data only}*

Estlander AM, Takala EP, Viikari-Juntura E. Do psychological factors predict changes in musculoskeletal pain? A prospective, two-year follow-up study of a working population. *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine* 1998;**40**(5):445-53.

Feuerstein 2006 *{published data only}*

Feuerstein M, Baldwin Harrington C, Lopez M, Haufler A. How do job stress and ergonomic factors impact clinic visits in acute low back pain? A prospective study. *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine* 2006;**48**(6):607-14.

Fitzpatrick 1987 *{published data only}*

Fitzpatrick RM, Bury M, Frank AO, Donnelly T. Problems in the assessment of outcome in a back pain clinic. *International Disability Studies* 1987;**9**(4):161-5.

Goossens 2005 *{published data only}*

Goossens ME, Vlaeyen JW, Hidding A, Kole-Snijders A, Evers SM. Treatment expectancy affects the outcome of cognitive-behavioral interventions in chronic pain. *Clinical Journal of Pain* 2005;**21**(1):18-26.

Hurwitz 2005 *{published data only}*

Hurwitz EL, Morgenstern H, Chiao C. Effects of recreational physical activity and back exercises on low back pain and psychological distress: Findings from the UCLA low back pain study. *American Journal of Public Health* 2005;**95**(10):1817-24.

Jellema 2007 *{published data only}*

Jellema P, Van der Windt DA, Van der Horst HE, Stalman WA, Bouter LM. Prediction of an unfavourable course of low back pain in general practice: comparison of four instruments. *British Journal of General Practice* 2007;**57**(534):15-22.

Kendell 2018 *{published data only}*

Kendell M, Beales D, O'Sullivan P, Rabey M, Hill J, Smith A. The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *Journal of Physiotherapy* 2018;**64**(2):107-13.

Lochting 2017 *{published data only}*

Lochting I, Garratt AM, Storheim K, Werner EL, Grotle m. The impact of psychological factors on condition-specific, generic and individualized patient reported outcomes in low back pain. *Health and Quality of Life Outcomes* 2017;**15**(1):40.

Lotters 2006 *{published data only}*

Lotters F, Burdorf A. Prognostic factors for duration of sickness absence due to musculoskeletal disorders. *Clinical Journal of Pain* 2006;**22**(2):212-21.

Lurie 2016 *{published data only}*

Lurie JD, Henderson ER, McDonough CM, Berven SH, Scherer EA, Tosteson TD, et al. Effect of expectations on treatment outcome for lumbar intervertebral disc herniation. *Spine* 2016;**41**(9):803-9.

Maxwell 1998 *{published data only}*

Maxwell TD, Gatchel RJ, Mayer TG. Cognitive predictors of depression in chronic low back pain: Toward an inclusive model. *Journal of Behavioural Medicine* 1998;**21**(2):131-43.

Melloh 2011 *{published data only}*

Melloh M, Elfering A, Presland CE, Roeder C, Hendrick P, Darlow B, et al. Predicting the transition from acute to persistent low back pain. *Occupational Medicine* 2011;**61**(2):127-31.

Ng 2017 *{published data only}*

Ng SK, Cicuttini FM, Wang Y, Wluka AE, Fitzgibbon B, Urquhart DM. Negative beliefs about low back pain are associated with persistent high intensity low back pain. *Psychology, Health & Medicine* 2017;**22**(7):790-9.

Reis 2007 *{published data only}*

Reis S, Borkan J, Van Raalte R, Tamir A, Dahan R, Hermoni D. The LBP patient perception scale: a new predictor of LBP episode outcomes among primary care patients. *Patient Education and Counseling* 2007;**67**(1-2):191-5.

Roberts 2015 *{published data only}*

Roberts AJ, Dew A, Bridger R, Etherington J, Kilminster S. Predicting low back pain outcome following rehabilitation for low back pain. *Journal of Back and Musculoskeletal Rehabilitation* 2015;**28**(1):119-28.

Schultz 2008 *{published data only}*

Schultz IZ, Crook J, Berkowitz J, Milner R, Meloche GR, Lewis ML. A prospective study of the effectiveness of early intervention with high-risk back-injured workers -- a pilot study. *Journal of Occupational Rehabilitation* 2008;**18**(2):140-51.

Silvis 2016 *{published data only}*

Silvis WL, Lakke SE, Stegeman P, Speijer BL, Vroomen PC, Coppes MH, et al. Can patients with low back pain be satisfied with less than expected?. *Physiotherapy* 2016;**102**:e54.

Skargren 1998 *{published data only}*

Skargren EI, Oberg BE. Predictive factors for 1-year outcome of low-back and neck pain in patients treated in primary care: comparison between the treatment strategies chiropractic and physiotherapy. *Pain* 1998;**77**(2):201-7.

Smeets 2008 *{published data only}*

Smeets RJ, Beelen S, Goossens ME, Schouten EG, Knottnerus JA, Vlaeyen JW. Treatment expectancy and credibility are associated with the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. *Clinical Journal of Pain* 2008;**24**(4):305-15.

Soucy 2006 *{published data only}*

Soucy I, Truchon M. Work-related factors contributing to chronic disability in low back pain. *Work (Reading, Mass.)* 2006;**26**(3):313-26.

Staerkle 2004 *{published data only}*

Staerkle R, Mannion AF, Elfering A, Junge A, Semmer NK, Jacobshagen N, et al. Longitudinal validation of the Fear-Avoidance Beliefs Questionnaire (FABQ) in a Swiss-German sample of low back pain patients. *European Spine Journal* 2004;**13**(4):332-40.

Steenstra 2015 *{published data only}*

Steenstra IA, Busse JW, Tolusso D, Davilmar A, Lee H, Furlan AD, et al. Predicting time on prolonged benefits for injured workers with acute back pain. *Journal of Occupational Rehabilitation* 2015;**25**(2):267-78.

Vargas-Prada 2013 *{published data only}*

Vargas-Prada S, Martínez JA, Coggon D, Delclos G, Benavides FG, Serra C. Health beliefs, low mood, and somatizing tendency: Contribution to incidence and persistence of musculoskeletal pain with and without reported disability. *Scandinavian Journal of Work, Environment and Health* 2013;**39**(6):589-98.

Wolff 2018 *{published data only}*

Wolff SV, Willburger R, Hallner D, Rusu AC, Rusche H, Schulte T, et al. Avoidance-endurance fast screening (AE-FS). Content and predictive validity of a 9-item screening instrument for patients with unspecific subacute low back pain. *Schmerz* 2018;**32**(4):283-92.

References to studies awaiting assessment

Ailliet 2018 *{published data only}*

Ailliet L, Rubinstein SM, Hoekstra T, Van Tulder MW, Vet HC. Long-term trajectories of patients with neck pain and low back pain presenting to chiropractic care: A latent class growth analysis. *European Journal of Pain* 2018;**22**(1):103-13.

Ashworth 2013 *{published data only}*

Ashworth J, Green DJ, Dunn KM, Jordan KP. Opioid use among low back pain patients in primary care: Is opioid prescription associated with disability at 6-month follow-up?. *Pain* 2013;**154**(7):1038-44.

Barons 2014 *{published data only}*

Barons MJ, Griffiths FE, Parsons N, Alba A, Thorogood M, Medley G F, et al. Matching patients to an intervention for back pain: classifying patients using a latent class approach. *Journal of Evaluation in Clinical Practice* 2014;**20**(4):544-50.

Friedman 2018 *{published data only}*

Friedman BW, Conway J, Campbell C, Bijur PE, Gallagher EJ, Bird SB. Pain one week after an Emergency Department visit for acute low back pain is associated with poor three-month outcomes. *Academic Emergency Medicine* 2018;**25**(10):1138-45.

Ganesh 2019 *{published data only}*

Ganesh GS, Sahu PK, Das SP, Mishra C, Dhiman S. A subgroup analysis to compare patients with acute low back pain classified as per treatment-based classification. *Physiotherapy Research International* 2019;**24**(1):e1747.

Glattacker 2018 *{published data only}*

Glattacker M, Heyduck K, Jakob T. Yellow flags as predictors of rehabilitation outcome in chronic low back pain. *Rehabilitation Psychology* 2018;**63**(3):408-17.

Harter 2004 *{published data only}*

Harter WH, Hoch FS, Denner A, Uhlig H. Training therapy in patients with chronic back pain - Development of a risk factor model for the prediction of compliance. *Zeitschrift fur Orthopadie und ihre Grenzgebiete* 2004;**142**(6):639-43.

Hartvigsen 2018a *{published data only}*

Hartvigsen L, Kongsted A, Vach W, Salmi L, Hestbaek L. Does a diagnostic classification algorithm help to predict the course of low back pain? A study of Danish chiropractic patients with 1-year follow-up. *Journal of Orthopaedic and Sports Physical Therapy* 2018;**48**(11):837-46.

Klyne 2018 *{published data only}*

Klyne DM, Barbe MF, Van den Hoorn W, Hodges PW. Longitudinal analysis of inflammatory, psychological, and sleep-related factors following an acute low back pain episode—the good, the bad, and the ugly. *European Spine Journal* 2018;**27**(4):763-77.

Mehling 2015 *{published data only}*

Mehling WE, Ebell MH, Avins AL, Hecht FM. Clinical decision rule for primary care patient with acute low back pain at risk of developing chronic pain. *Spine Journal* 2015;**15**(7):1577-86.

Mendelson 1983 *{published data only}*

Mendelson G, Selwood TS, Kranz H, Loh TS, Kidson MA, Scott DS. Acupuncture treatment of chronic back pain. A double-blind placebo-controlled trial. *American Journal of Medicine* 1983;**74**(1):49-55.

Pfingsten 1997 *{published data only}*

Pfingsten M, Hildebrandt J, Leibing E, Franz C, Saur P. Effectiveness of a multimodal treatment program for chronic low-back pain. *Pain* 1997;**73**(1):77-85.

Pfingsten 1997a *{published data only}*

Pfingsten M, Hildebrandt J, Saur P, Franz C, Seeger D. Multidisciplinary treatment program on chronic low back pain, part 4. Prognosis of treatment outcome and final conclusions. *Schmerz (Berlin, Germany)* 1997;**11**(1):30-41.

Pfingsten 1997b *{published data only}*

Pfingsten M, Leibing E, Franz C, Bansemer D, Busch O, Hildebrandt J. Fear-avoidance-beliefs in patients with backpain. *Schmerz (Berlin, Germany)* 1997;**11**(6):387-95.

Thomas 2005 *{published data only}*

Thomas KJ, MacPherson H, Ratcliffe J, Thorpe L, Brazier J, Campbell M, et al. Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain. *Health Technology Assessments* 2005;**9**(32):iii-iv, ix-x, 1-109.

Weber 1998 *{published data only}*

A Weber, M Wilhelm, U Weber, H Raspe. Is the subjective health status a good predictor of work resumption?. *Sozial-Und Praventivmedizin* 1998;**43**(4):177-84.

Additional references

Anderson 1995

Anderson KO, Dowds BN, Pelletz RE, Edwards WT, Peeters-Asdourian C. Development and initial validation of a scale to measure self-efficacy beliefs in patients with chronic pain. *Pain* 1995;**63**(1):77-84. [PUBMED: 8577493]

Bandura 1977

Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review* 1977;**84**(2):191-215.

Bandura 2004

Bandura A. Health promotion by social cognitive means. *Health Education & Behavior* 2004;**31**(2):143-64.

Borenstein 2009

Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. West Sussex: John Wiley & Sons, Ltd, 2009.

Busse 2019

Busse JW, Heels-Ansdell D, Makosso-Kallyth S, Petrisor B, Jeray K, Tufescu T, et al. Patient coping and expectations predict recovery after major orthopaedic trauma. *British Journal of Anaesthesia* 2019;**122**(1):51-9. [PUBMED: 30579406]

Darlow 2012

Darlow B, Fullen BM, Dean S, Hurley DA, Baxter GD, Dowell A. The association between health care professional attitudes and beliefs and the attitudes and beliefs, clinical management, and outcomes of patients with low back pain: A systematic review. *European Journal of Pain* 2012;**16**(1):3-17.

Dawes 2016

Dawes AJ, Maggard-Gibbons M, Maher AR, Booth MJ, Miake-Lye I, Beroes JM, et al. Mental health conditions among patients seeking and undergoing bariatric surgery: a meta-analysis. *JAMA* 2016;**315**(2):150-63.

Devilly 2000

Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry* 2000;**31**(2):73-86.

Dionne 2008

Dionne CE, Dunn KM, Croft PR, Nachemson AL, Buchbinder R, Walker BF, et al. A consensus approach toward the standardization of back pain definitions for use in prevalence studies. *Spine (Phila Pa 1976)* 2008;**33**(1):95-103.

DistillerSR [Computer program]

Evidence Partners. DistillerSR. Ottawa, ON: Evidence Partners, 2018.

Ebrahim 2015

Ebrahim S, Malachowski C, Kamal El Din M, Mulla SM, Montoya L, Bance S, et al. Measures of patients' expectations about recovery: a systematic review. *Journal of Occupational Rehabilitation* 2015;**25**(1):240-55. [PUBMED: 25100443]

Egger 1997

Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 1997;**315**:629-34.

EndNote X8 [Computer program]

Clarivate Analytics. EndNote X8. Clarivate Analytics, 2018.

EuroQol Research Foundation 2019

EuroQol Research Foundation. EQ-5D-5L User Guide. https://euroqol.org/publications/user-guides 2019.

Fadyl 2008

Fadyl J, McPherson K. Return to work after injury: a review of evidence regarding expectations and injury perceptions, and their influence on outcome. *Journal of Occupational Rehabilitation* 2008;**18**(4):362-74.

Fairbank 1980

Fairbank J, Couper J, Davies J, O'Brien JP. The Oswestry low back pain questionnaire. *Physiotherapy* 1980;**66**:271-3.

Freburger 2009

Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, Wallace AS, et al. The rising prevalence of chronic low back pain. *Archives of Internal Medicine* 2009;**169**(3):251-8.

Furlan 2015

Furlan AD, Malmivaara A, Chou R, Maher CG, Deyo RA, Schoene M, et al. Updated method guideline for systematic reviews in the Cochrane Back and Neck Group. *Spine* 2015;**40**(21):1660-73.

GBD Collaborators 2016

GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016;**388**(10053):1545-602.

Gross 2005

Gross DP, Battie MC. Work-related recovery expectations and the prognosis of chronic low back pain within a workers' compensation setting. *Journal of Occupational and Environmental Medicine* 2005;**47**(4):428-33.

Guyatt 2011

Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *Journal of Clinical Epidemiology* 2011;**64**(4):380-2.

Guzman 2007

Guzman J, Hayden J, Furlan AD, Cassidy JD, Loisel P, Flannery J, et al. Key factors in back disability prevention: a consensus panel on their impact and modifiability. *Spine (Phila Pa 1976)* 2007;**32**(7):807-15.

Hallegraeff 2012

Hallegraeff JM, Krijnen WP, Van der Schans CP, De Greef MH. Expectations about recovery from acute non-specific low back pain predict absence from usual work due to chronic low back pain: a systematic review. *Journal of Physiotherapy* 2012;**58**(3):165-72.

Hartvigsen 2004

Hartvigsen J, Lings S, Leboeuf-Yde C, Bakketeig L. Psychosocial factors at work in relation to low back pain and consequences of low back pain; a systematic, critical review of prospective cohort studies. *Occupational and Environmental Medicine* 2004;**61**(1):e2.

Hartvigsen 2018b

Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. *Lancet* 2018;**391**(10137):2356-67.

Haskins 2012

Haskins R, Rivett DA, Osmotherly PG. Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. *Manual Therapy* 2012;**17**(1):9-21.

Hayden 2007

Hayden JA. Methodological Issues in Systematic Reviews of Prognosis and Prognostic Factors: Low Back Pain [Doctoral Thesis]. Toronto, Canada: University of Toronto, 2007.

Hayden 2008

Hayden JA, Cote P, Steenstra IA, Bombardier C. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. *Journal of Clinical Epidemiology* 2008;**61**(6):552-60.

Hayden 2009

Hayden JA, Chou R, Hogg-Johnson S, Bombardier C. Systematic reviews of low back pain prognosis had variable methods and results: guidance for future prognosis reviews. *Journal of Clinical Epidemiology* 2009;**62**(8):781-96 e1.

Hayden 2010

Hayden JA, Dunn KM, Van der Windt DA, Shaw WS. What is the prognosis of back pain?. *Best Practice & Research. Clinical Rheumatology* 2010;**24**(2):167-79.

Hayden 2013

Hayden JA, Van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. *Annals of Internal Medicine* 2013;**158**(4):280-6.

Hemingway 1999

Hemingway H, Marmot M. Evidence based cardiology: psychosocial factors in the aetiology and prognosis of coronary heart disease. Systematic review of prospective cohort studies. *BMJ (Clinical research ed.)* 1999;**318**(7196):1460-7.

Hendrick 2011

Hendrick P, Milosavljevic S, Hale L, Hurley DA, McDonough S, Ryan B, et al. The relationship between physical activity and low back pain outcomes: a systematic review of observational studies. *European Spine Journal* 2011;**20**(3):464-74.

Hestbaek 2003

Hestbaek L, Leboeuf-Yde C, Manniche C. A review of studies of general patient populations. *European Spine Journal* 2003;**12**(2):149-65.

Holmes 2016

Holmes SD, Fornaresio LM, Miller CE, Shuman DJ, Ad N. Development of the Cardiac Surgery Patient Expectations Questionnaire (C-SPEQ). *Quality of life Research* 2016;**25**(8):2077-86. [PUBMED: 26883817]

Hoy 2010

Hoy D, Brooks P, Blyth F, Buchbinder R. The epidemiology of low back pain. *Best Practice & Research. Clinical Rheumatology* 2010;**24**(6):769-81.

Huguet 2013

Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. *Systematic Reviews* 2013;**2**:71. [DOI: 10.1186/2046-4053-2-71]

Iles 2008

Iles RA, Davidson M, Taylor NF. Psychosocial predictors of failure to return to work in non-chronic non-specific low back pain: a systematic review. *Occupational and Environmental Medicine* 2008;**65**(8):507-17.

Iles 2009

Iles RA, Davidson M, Taylor NF, O'Halloran P. Systematic review of the ability of recovery expectations to predict outcomes in non-chronic non-specific low back pain. *Journal of Occupational Rehabilitation* 2009;**19**(1):25-40.

Iles 2011

Iles RA, Taylor NF, Davidson M, O'Halloran P. Telephone coaching can increase activity levels for people with non-chronic low back pain: a randomized trial. *Journal of Physiotherapy* 2011;**57**:231-8.

Jimenez 2009

Jimenez D, Uresandi F, Otero R, Lobo JL, Monreal M, Marti D, et al. Troponin-based risk stratification of patients with acute nonmassive pulmonary embolism: systematic review and metaanalysis. *Chest* 2009;**136**(4):974-82.

Kapoor 2006

Kapoor S, Shaw WS, Pransky G, Patterson W. Initial patient and clinician expectations of return to work after acute onset of work-related low back pain. *Journal of Occupational and Environmental Medicine* 2006;**48**(11):1173-80.

Kovacs 2007

Kovacs FM, Abraira V, Royuela A, Corcoll J, Alegre L, Cano A, et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine (Phila Pa 1976)* 2007;**32**(25):2915-20.

Lamberink 2017

Lamberink HJ, Otte WM, Geerts AT, Pavlovic M, Ramos-Lizana J, Marson AG, et al. Individualised prediction model of seizure recurrence and long-term outcomes after withdrawal of antiepileptic drugs in seizure-free patients: a systematic review and individual participant data meta-analysis. *Lancet* 2017;**16**(7):523-31.

Levin 1996

Levin JB, Lofland KR, Cassisi JE, Poreh AM, Blonsky ER. The relationship between self-efficacy and disability in chronic low back pain patients. *International Journal of Rehabilitation and Health* 1996;**2**(1):19-28.

Lim 2007

Lim HS, Chen PP, Wong TC, Gin T, Wong E, Chan IS, et al. Validation of the Chinese version of pain self-efficacy questionnaire. *Anesthesia and Analgesia* 2007;**104**(4):918-23.

Lim 2012

Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;**380**(9859):2224-60.

Linton 1998

Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clinical Journal of Pain* 1998;**14**(3):209-15. [PUBMED: 9758070]

Melzack 1975

Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods. *Pain* 1975;**1**:277-99.

Menezes Costa 2012

Menezes Costa LC, Maher CG, Hancock MJ, McAuley JH, Herbert RD, Costa LO. The prognosis of acute and persistent low-back pain: a meta-analysis. *CMAJ : Canadian Medical Association Journal* 2012;**184**(11):613-24.

Metcalfe 2005

Metcalfe C, Moffett JK. Do patients' expectations of physiotherapy affect treatment outcome? Part 1: Baseline data. *International Journal of Therapy and Rehabilitation* 2005;**12**(2):5-62.

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology* 2009;**62**(10):1006-12.

Moss-Morris 2002

Moss-Morris R, Weinman J, Petrie KJ, Horne R, Cameron LD, Buick D. The revised illness perception questionnaire (IPQ-R). *Psychology and Health* 2002;**17**(1):1-16.

Nicholas 2007

Nicholas MK. The pain self-efficacy questionnaire: taking pain into account. *European Journal of Pain* 2007;**11**(2):153-63.

Ostelo 2008

Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 2008;**33**(1):90-4.

Parsons 2007

Parsons S, Harding G, Breen A, Foster N, Pincus T, Vogel S, et al. The influence of patients' and primary care practitioners' beliefs and expectations about chronic musculoskeletal pain on the process of care: a systematic review of qualitative studies. *Clinical Journal of Pain* 2007;**23**(1):91-8.

Peat 2014

Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KG, et al. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. *PLoS Medicine* 2014;**11**(7):e1001671. [PUBMED: 25003600]

Peterson 2005

Peterson RA, Brown SP. On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology* 2005;**90**(1):175-81.

Price 1999

Price D, Barrell J. Expectation and desire in pain and pain reduction. In: Kirsch I editor(s). How Expectancies Shape Experience. Washington, DC: American Psychological Association, 1999.

Radloff 1977

Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977;**1**:385-401.

Ramond 2011

Ramond A, Bouton C, Richard I, Roquelaure Y, Baufreton C, Legrand E, et al. Psychosocial risk factors for chronic low back pain in primary care--a systematic review. *Family Practice* 2011;**28**(1):12-21.

Riley 2013

Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Medicine* 2013;**10**(2):e1001380.

Riley 2019

Riley RD, Moons KG, Snell KI, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ (Clinical Research ed.)* 2019;**364**:k4597. [PUBMED: 30700442]

Robins 2001

Robins R, Hendin HM, Trzesniewski K. Measuring global self-esteem: construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin* 2001;**27**(2):151-61.

Roland 2000

Roland M, Fairbank J. The Roland Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine* 2000;**25**:3115-24.

Sarda 2007

Sarda J Jr, Nicholas MK, Pimenta CA, Asgharl A. Pain-related self-efficacy beliefs in a Brazilian chronic pain patient sample: a psychometric analysis. *Stress and Health* 2007;**23**(3):185-90.

Schultz 2002

Schultz IZ, Crook JM, Berkowitz J, Meloche GR, Milner R, Zuberbier OA, et al. Biopsychosocial multivariate predictive model of occupational low back disability. *Spine* 2002;**27**(23):2720-25.

Schultz 2005

Schultz IZ, Crook J, Berkowitz J, Milner R, Meloche GR. Predicting return to work after low back injury using the psychosocial risk for occupational disability instrument: a validation study. *Journal of Occupational Rehabilitation* 2005;**15**(3):365-76.

Shaw 2011

Shaw WS, Reme SE, Linton SJ, Huang YH, Pransky G. 3rd place, PREMUS best paper competition: development of the return-to-work self-efficacy (RTWSE-19) questionnaire--psychometric properties and predictive validity. *Scandinavian Journal of Work, Environment & Health* 2011;**37**(2):109-19. [PUBMED: 21203742]

Steenstra 2012

Steenstra IA, Lee H, De Vroome EM, Busse JW, Hogg-Johnson SJ. Comparing current definitions of return to work: a measurement approach. *Journal of Occupational Rehabilitation* 2012;**22**(3):394-400.

Symons 2002

Symons MJ, Moore DT. Hazard rate ratio and prospective epidemiological studies. *Journal of Clinical Epidemiology* 2002;**55**(9):893-9.

Tate 1999

Tate RB, Yassi A, Cooper J. Predictors of time loss after back injury in nurses. *Spine (Phila Pa 1976)* 1999;**24**(18):1930-5; discussion 36.

Vlaeyen 1990

Vlaeyen JW, Geurts SM, Kole-Snijders AM, Schuerman JA, Groenman NH, Van Eek H. What do chronic pain patients think of their pain? Towards a pain cognition questionnaire. *British Journal of Clinical Psychology* 1990;**29**(Pt 4):383-94. [PUBMED: 2149663]

Vlaeyen 2000

Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain* 2000;**85**(3):317-32.

Ware 1992

Ware JE, Sherbourne CD. The MOS 36 item short-form health survey (SF-36). 1: Conceptual framework and item selection. *Med Care* 1992;**30**:473-80.

WHO 2002

WHO (World Health Organization). The International Classification of Functioning, Disability and Health (ICF). Geneva: WHO, 2002.

Wilczynski 2004

Wilczynski NL, Haynes RB, the Hedges team. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Medicine* 2004;**9**(2):23.

References to other published versions of this review

Hayden 2014

Hayden JA, Tougas ME, Riley R, Iles R, Pincus T. Individual recovery expectations and prognosis of outcomes in non-specific low back pain: Prognostic factor exemplar review. *Cochrane Database of Systematic Reviews* 2014, Issue 9. [DOI: 10.1002/14651858.CD011284]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Beneciuk 2017

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People consulting general practice for back pain; United Kingdom; Primary care; Data collection period not reported
Sample size	851
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectations of recovery (0 - 10)
Notes	Funding Source: The Arthritis Research United Kingdom; West Midlands North Comprehensive Local Research Network; North Staffordshire Primary Care Research Consortium Conflict of Interest: None declared [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Besen 2015

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Volunteers seeking treatment for work-related, acute back pain at private occupational medicine clinics; United States of America; Occupational; Data collection period not reported
Sample size	496
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: How soon do you expect to be able to resume your normal job without any limitations? (6 pts) SELF-EFFICACY: Return-To-Work Self-Efficacy Scale (19 items, 1 - 10)
Notes	Funding Source: Not reported Conflict of Interest: Not reported [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Bishop 2015

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People recruited from primary care research network, acupuncturist associations, and internet searches for hospital-based services; United Kingdom; Secondary care; November 2008 - October 2010
Sample size	420
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: Credibility Expectancy Questionnaire (3 items, 0 – 9) SELF-EFFICACY: Chronic Pain Self-Efficacy for Pain Management subscale (0 - 100)

Bishop 2015 *(Continued)*

Notes	Funding Source: Arthritis Research United Kingdom Conflict of Interest: None declared

Butler 2007

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Workers who filed claims for occupational back pain (with or without leg pain or sciatica); United States of America; Occupational; January 1st 1999 - June 30th 2002
Sample size	1831
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectations of recovery (5 pts)
Notes	Funding Source: Not reported Conflict of Interest: Not reported

Carriere 2015

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Consecutive referrals for work-related back pain (primary diagnosis) to primary care physical therapy clinics; Canada; Primary care; Data collection period not reported
Sample size	109
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Likelihood of return to work (0 - 100)
Notes	Funding Source: Canadian Institute for Health Research; Institut de recherche Robert-Sauve en sante et en securite du travail Conflict of Interest: Not reported

Casey 2008

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People presenting at an acute back pain clinic; United States of America; Secondary care; Data collection period not reported
Sample size	84
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Pain Behaviour and Perception Inventory pain permanence subscale (4 pts)

Casey 2008 *(Continued)*

Notes	Funding Source: Not reported Conflict of Interest: Not reported

Demmelmaier 2010

Population (low back pain description; location; setting; data collection period)	Prospective cohort: 2 back pain groups, 1 with first-episode pain, and 1 with long-term pain; Sweden; General population; February 2005 - February 2006
Sample size	379
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL (2): Pain expectations (2 items, 7 pts); Work expectations in 6 months (0 - 6)
Notes	Funding Source: The Olle Engkvist Building Foundation, Sweden Conflict of Interest: Not reported

Dionne 1997

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People consulting a primary care physician for back pain; United States of America; Primary care; 1989 - 1990
Sample size	1213
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectation of continued pain (5 pts)
Notes	Funding Source: Agency for Health Care Policy and Research; The National Institute for Dental Research; National Health Research and Development Program of Canada Conflict of Interest: Not reported [a]Study does not present sufficient data to be included in meta-analyses (univariate), but does provide some measure of statistical significance

Downie 2016

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People seeking primary care for an episode of acute low back pain; Australia; Primary care; Data collection period not reported
Sample size	653

Downie 2016 *(Continued)*

Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Risk of persistence (0 - 10)
Notes	Funding Source: "No specific grant from any funding agency in the public, commerical, or not-for-profit sectors" Conflict of Interest: None declared [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Du Bois 2008

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Workers entitled to sickness allowance; Belgium; Occupational; Data collection period not reported
Sample size	186
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Return to work within 6 months (0 - 10)
Notes	Funding Source: Not reported Conflict of Interest: Not reported

Enthoven 2006

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People with low back pain who were eligible for sick-leave benefits; Sweden; Primary care; Data collection period not reported
Sample size	148
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: Expectations of treatment (5 pt)
Notes	Funding Source: Not reported Conflict of Interest: Not reported [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Enthoven 2016

Population (low back pain description; location; set-	Prospective cohort: People consulting their general practitioner with a new episode of back pain; Netherlands; Primary care; Data collection period not reported

Enthoven 2016 *(Continued)*

ting; data collection period)	
Sample size	675
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectations of recovery in 3 months (1 - 5)
Notes	Funding Source: Department of General Practice, Erasmus University Medical Center, Rotterdam, Coolsingel Foundation, Rotterdam, and the Dutch Arthritis Foundation Conflict of Interest: Not reported

Foster 2008

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Consecutive patients presenting to general practices with low back pain; United Kingdom; Primary care; September 2004 - April 2006
Sample size	1591
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Revised Illness Perceptions Questionnaire, timeline acute/chronic item (5 pts) SELF-EFFICACY: Pain Self-Efficacy Questionnaire (10 items, 6 pts)
Notes	Funding Source: Arthritis Research Campaign, United Kingdom; North Staffordshire Primary Care Research Consortium Conflict of Interest: Not reported [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

George 2010

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: Consecutive patients seeking treatment for low back pain at university health clinics; United States of America; Primary care; Data collection period not reported
Sample size	105
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectations of recovery (1 - 5)
Notes	Funding Source: National Institutes of Health-National Institute of Arthritis and Musculoskeletal and Skin Diseases Grant Conflict of Interest: None declared

Gervais 1991

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Male workers receiving workers' compensation reporting pain from thoracic (T9) to sacrum (S1) spine level; Canada; Occupational; Data collection period not reported
Sample size	135
Prognostic factor (type, measure(s) of individual recovery expectations)	SELF-EFFICACY: Self-Efficacy and Results Expectancies Inventory (unclear scale)
Notes	Funding Source: Comité d'attribution de fonds internes de recherche Research Fund of the University of Montreal Conflict of Interest: Not reported

Glattacker 2013

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial (control group only): People with chronic low back pain recruited from orthopedic rehabilitation centres; Germany; Secondary care; December 2008 - April 2010
Sample size	105
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Revised Illness Perceptions Questionnaire, timeline acute/chronic item (5 pts)
Notes	Funding Source: German Pension Insurance Conflict of Interest: Not reported

Goldstein 2002

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People who sought care for low back pain from their health maintenance organization provider; United States of America; Primary care; October 30th 1995 - November 9th 1998
Sample size	681
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: Treatment confidence, Numeric Rating Scale (0 - 10)
Notes	Funding Source: Not reported Conflict of Interest: Not reported

Gross 2010

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Workers' compensation claimants with back pain seen at a rehabilitation facility for return to work assessment; Canada; Secondary care; April 1st 2001 - March 31st 2002
Sample size	298
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Work-Related Recovery Expectations Questionnaire (3 items, 5 pts)
Notes	Funding Source: Workers' Compensation Board of British Columbia; Workers' Compensation Board of Alberta Conflict of Interest: Not reported

Grotle 2006

Population (low back pain description; location; setting; data collection period)	Prospective cohort: 2 samples, 1 with acute first-time low back pain contacting primary health care, 1 with chronic low back pain recruited from a specialist back clinic; Norway; Mixed population; Data collection period not reported
Sample size	173
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL (2): Risk of persistent pain (0 - 10), and Certainty of working in 6 months (0 - 10)
Notes	Funding Source: Not reported Conflict of Interest: Not reported

Haas 2014

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People with chronic low back pain recruited from general population; United States of America; General population; March 2007 - May 2010
Sample size	391
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: Confidence in treatment (2 items, 6 pts)
Notes	Funding Source: The National Center for Complementary and Alternative Medicine; National Institutes of Health Conflict of Interest: None declared [a]Study does not present sufficient data to be included in meta-analyses (univariate), but does provide some measure of statistical significance

Hagen 2005

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People sick-listed 8 to 12 weeks for low back pain (with or without radiating pain); Norway; Primary care; Data collection period not reported
Sample size	457
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Belief that back pain will disappear from Graded Reduced Work Ability Scale (1 - 6)
Notes	Funding Source: Norwegian Ministry of Health and Social Affairs Conflict of Interest: Not reported

Haldorsen 1998

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People who were employed and on sick leave for low back pain; Norway; Occupational; Data collection period not reported
Sample size	223
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL (2): Do you believe that you will be back to work after a couple of weeks? (5 pts), and Do you believe your complaints will be less during the first couple of weeks? (5 pts) SELF-EFFICACY: If you continue working, what effect will that have on your complaints? (5 pt)
Notes	Funding Source: The Royal Norwegian Department of Health and Social Affairs Conflict of Interest: None declared

Harkapaa 1996

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People with chronic or recurrent low back pain referred to rehabilitation and receiving disability payments; Finland; Secondary care; Data collection period not reported
Sample size	175
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Health Optimism Index (5 - 20)
Notes	Funding Source: Not reported Conflict of Interest: Not reported [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Hazard 1996

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Workers reporting occupational low back injury to the Department of Labor and Industry within 11 days of onset; United States of America; Occupational; September 1993 - June 1994
Sample size	166
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Certainty of working in 6 months (0 - 10)
Notes	Funding Source: National Institute on Disability and Rehabilitation Research, United States Department of Education Conflict of Interest: Not reported

Henschke 2008

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: Consecutive patients with non-specific low back pain < 2 weeks duration recruited from the clinics of general practitioners, physiotherapists, and chiropractors; Australia; Primary care; November 2003 - July 2005
Sample size	969
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Perceived risk of persistence (0 - 10)
Notes	Funding Source: National Health and Medical Research Council Conflict of Interest: None declared [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Heymans 2006

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People who visited their occupational physician on sick leave for < 8 weeks; Netherlands; Occupational; October 2000 - November 2002
Sample size	299
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL (2): When do you think you will be able to work fulltime again? (8 pts), and How certain are you about full work resumption at 6 months (5 pts); TREATMENT: Expected benefit from treatment (0 - 10)
Notes	Funding Source: The Netherlands Organization for Health Research and Development; Dutch Minisitry of Health, Welfare and Sport; Dutch Ministry of Social Affairs and Employment Conflict of Interest: Not reported

Hildebrandt 1997

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People recruited from a pain clinic with chronic back pain and at least 3 months of disability leave during the last year; Germany; Secondary care; Data collection period not reported
Sample size	90
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectation of ability to return to work after discharge from treatment (unclear scale)
Notes	Funding Source: The Germany Ministry of Education, Research, and Technology Conflict of Interest: Not reported

Jellema 2002

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Employees of a home care organisation with current low back pain or ≥ 2 episodes in past year; Netherlands; Occupational; Data collection period not reported
Sample size	59
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT (2): Confidence in reduction in pain by lumbar support (0 - 10), and Confidence in improvement in function by lumbar support (0 - 10)
Notes	Funding Source: The Netherlands Organization for Health Research and Development Conflict of Interest: Not reported

Jensen 2000

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial and Prospective cohort: Employees of a homecare organisation with low back pain; Sweden; Secondary care; August 1995 - September 1998
Sample size	235
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: Belief that there is a treatment that could relieve condition (unclear scale) SELF-EFFICACY: Belief in ability for learning to cope with the pain (unclear scale)
Notes	Funding Source: AMF Insurance; SPP Insurance Conflict of Interest: Not reported

Jensen 2013

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People sick-listed for low back pain (with or without radiculopathy); Denmark; Secondary care; 2004 - 2009
Sample size	325
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectations about return to work within 6 months (10 pts)
Notes	Funding Source: Municipality of Silkeborg; Municipality of Favrskov; Municipality of Skanderborg; Municipality of Denmark; The Central Denmark Region; The Danisk Working Environment Research Fund Conflict of Interest: None declared

Karjalainen 2003

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People with disabling low back pain for the preceding 4 to 12 weeks; Finland; Primary care; Data collection period not reported
Sample size	164
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Perceived risk of not recovering (0 – 10) TREATMENT: Expectations about the effectiveness of treatment (0 – 10)
Notes	Funding Source: Not reported Conflict of Interest: Not reported [a]Study does not present sufficient data to be included in meta-analyses (univariate), but does provide some measure of statistical significance

Kongsted 2014

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People with low back pain consulting clinics in a research network; Denmark; Secondary care; September 2010 - January 2012
Sample size	928[a]
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Likelihood of recovery (0 - 10)
Notes	Funding Source: Danish Chiropractors' Foundation and Macroeconomic Policy Institute Almene Fond grant Conflict of Interest: None declared

Kongsted 2014 *(Continued)*

	[a]Kongsted 2014 was presented together, and separately as Kongsted 2014a (general practice cohort), and Kongsted 2014b (chiropractic practice cohort)

Leboeuf-Yde 2004

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Chiropractic patients who had recurrent low back pain with no chiropractic treatment in past 6 months ; Norway; Secondary care; Data collection period not reported
Sample size	843
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL (3): Will you be on sick leave in 6 weeks? (Yes/No/Don't know), Will you receive disability pension? (Yes/No/Don't know), and How will your low back pain be in 6 weeks? (5-pt scale) TREATMENT: Will chiropractic treatment help? (Yes/No/Don't know)
Notes	Funding Source: Research Council of Norway; the Swedish Chiropractic Association Conflict of Interest: Not reported

Lindell 2010

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: Primary care patients with non-acute non-specific spinal pain; Sweden; Primary care; August 2000 - January 2004
Sample size	125
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Self-prediction of probability of return to work at some time in the future (5 pts)
Notes	Funding Source: Stockholm County Social Insurance Agency; Stockholm County Council; Swedish Ministry of Health and Social Affairs; Vardal Foundation; Cardionics; Pharmacia; Grunenthal Sweden Conflict of Interest: None declared

Macedo 2014

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People from physical therapy clinics with chronic low back pain, assessed to be capable of physical exercise; Australia; Primary care; October 2007 - November 2009
Sample size	172
Prognostic factor (type, measure(s) of individual recovery expectations)	SELF-EFFICACY: Pain Self-Efficacy Questionnaire (0 - 100)
Notes	Funding Source: Australia's National Health and Medical Research Council

Macedo 2014 *(Continued)*

	Conflict of Interest: None declared

Magnussen 2007

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: Individuals receiving disability pension due to back pain for more than 1 year; Norway; Occupational; April 2004 - August 2005
Sample size	431
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Do you believe you will ever be able to return to work? (Yes/No/Don't know)
Notes	Funding Source: The Norwegian Foundation for Health and Rehabilitation Conflict of Interest: Not reported

Michaelson 2004

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People with chronic low back or neck pain referred to inpatient rehabilitation centre; Sweden; Secondary care; August 1997 - November 1999
Sample size	167
Prognostic factor (type, measure(s) of individual recovery expectations)	SELF-EFFICACY: Optimism Index (10 items)
Notes	Funding Source: VINNOVA; EG mal 1 Sapmi Conflict of Interest: Not reported [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes [b]Study does not present sufficient data to be included in meta-analyses (multivariate), but does provide some measure of statistical significance

Morlock 2002

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People in the physical therapy department of a multispecialty group practice; United States of America; Secondary care; Janaury 1999 - June 1999
Sample size	111

Morlock 2002 *(Continued)*

Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: Treatment expectations scale (5 items, 5 pts)
Notes	Funding Source: Health Alliance Plan of Michigan Conflict of Interest: Not reported

Myers 2007

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: Patients presenting for initial evaluation of low back pain and scored greater than 3 on a 0 – 10 pain scale; United States of America; Primary care; Data collection period not reported
Sample size	442
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: how much improvement do you expect in 6 weeks (0 - 10) TREATMENT: How helpful do you think the specified CAM therapy would be for your current episode of back pain or sciatica? (0 - 10)
Notes	Funding Source: National Center for Complementary and Alternative Medicine; Bernard Osher Foundation; American Specialty Health Conflict of Interest: None declared

Niemisto 2004

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: Employed people with low back pain (with or without sciatica) of at least 3 months' duration with a score of > 16% on the Oswestry Disability Index; Finland; General; Data collection period not reported
Sample size	204
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Self-rated prognosis for ability to work after 2 years (1 - 7)
Notes	Funding Source: Not reported Conflict of Interest: Not reported [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Opsahl 2016

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: Employees with long-lasting low back pain recruited from 4 clinics; Norway; Occupational; Data collection period not reported

Opsahl 2016 *(Continued)*

Sample size	286 (women); 283 (men)[a]
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: To what extent do you think you will return to work? (4 pts)
Notes	Funding Source: The Research Council of Norway Conflict of Interest: None declared [a]Opsahl 2016a (women) and Opsahl 2016b (men) are same study, presented separately

Opsommer 2017

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People with chronic low back pain admitted to a rehabilitation centre; Switzerland; Primary care; February 2011 - October 2013
Sample size	98
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL (2): In your view, how large is the risk that your current pain may become persistent? (0 - 10), and In your estimation, what are the chances that you will be able to work in six months? (0 - 10)
Notes	Funding Source: The Swiss National Science Foundation; HES-SO Conflict of Interest: None declared

Petersen 2007

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People referred to an outpatient rheumatology clinic for low back pain; Denmark; Secondary care; August 1996 - December 1998
Sample size	260
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Certainty of working 6 months after treatment (11 pts) SELF-EFFICACY: Expecting problems coping with future tasks (11 pts)
Notes	Funding Source: The Danish Physiotherapy Organization; Madsens Fund; The Danish Rheumatism Association Conflict of Interest: None declared [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Rasmussen-Barr 2012

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People with recurrent low back pain seeking care at a primary healthcare setting; Sweden; Primary care; Data collection period not reported
Sample size	71
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: Expectation of treatment (unclear scale)
Notes	Funding Source: Capio Research Foundation Conflict of Interest: None declared [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Reeser 2001

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People with acute low back pain from primary and tertiary care sites; United States of America; Mixed population; March 1999 - March 2000
Sample size	128
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: Expectations of treatment (5 items, 5 pts)
Notes	Funding Source: State Medical Society of Wisconsin Foundation; Pfizer, Inc.; Mercury Marine; Roche Pharmaceuticals; Deere & Company (Horicon); Monsanto Fund; Quad Graphics; Pharmacia Conflict of Interest: Not reported

Reiso 2003

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People with back disorders that primary healthcare doctors thought would certify them as sick for > 2 months; Norway; Primary care; September 1997 - December 1998
Sample size	190
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Self-predicted work status in 4 weeks (3 pts)
Notes	Funding Source: The Norweigan Ministry of Health and Social Affairs Conflict of Interest: None declared

Reme 2009

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People sick-listed for 8 - 12 weeks with low back pain; Norway; Mixed population; April 2000 - February 2004
Sample size	246
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectation to return to work within the next few weeks (Yes/No/No opinion)
Notes	Funding Source: Norwegian Foundation for Health and Rehabilitation Conflict of Interest: None declared

Rundell 2017

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Older adults (≥ 65 years old) presenting to primary care settings for new back pain visit, any duration back symptoms; United States of America; Primary care; 2011 - 2013
Sample size	5220
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectation for recovery (0 - 10)
Notes	Funding Source: The Agency for Healthcare Research and Quality Conflict of Interest: JGJ has served on the Comparative Effectiveness Advisory Board for General Electric Healthcare, is a cofounder and stockholder of PhysioSonics and receives royalties for intellectual property; also serves as a consultant for HealthHelp a radiology benefits management company [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes [b]Study does not present sufficient data to be included in meta-analyses (univariate), but does provide some measure of statistical significance

Sandstrom 1986

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People with non-specific low back pain referred to Department of Orthopaedic Surgery; Sweden; Secondary care; Data collection period not reported
Sample size	52
Prognostic factor (type, measure(s) of individual recovery expectations)	SELF-EFFICACY: I am afraid to start working again because I don't think I will be able to manage (7 pts)
Notes	Funding Source: The Goteborg Medical Society; The Asker Foundation; The Delegation for Social Research within the Swedish Ministry for Health and Social Affairs

Sandstrom 1986 *(Continued)*

	Conflict of Interest: Not reported [a]Study does not present sufficient data to be included in meta-analyses (multivariate), but does provide some measure of statistical significance

Schultz 2004

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Workers' Compensation Board low back injury claimants; Canada; Occupational; Data collection period not reported
Sample size	253
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectations of recovery scale (8 items, unclear scale)
Notes	Funding Source: Workers' Compensation Board of British Columbia; Workers' Compensation Board of Alberta Conflict of Interest: Not reported

Shaw 2009

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People seeking treatment at occupational health clinics for work-related, acute back pain; United States of America; Occupational; Data collection period not reported
Sample size	519
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expectation of fully returning to work in 4 weeks (3 pts)
Notes	Funding Source: Not reported Conflict of Interest: Not reported [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Sherman 2009

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People with chronic non-specific low back pain recruited from integrated healthcare systems; United States of America; Primary care; Data collection period not reported
Sample size	638

Sherman 2009 *(Continued)*

Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: Expectation of acupuncture helpfulness (0 - 10) SELF-EFFICACY: Likelihood of self-managing future back pain (unclear scale)
Notes	Funding Source: The National Center for Complementary and Alternative Medicine Conflict of Interest: Project officer for funder involved in analysis and interpretation of data and review and approval of manuscript

Steenstra 2005

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Hospital workers who reported sick leave due to nonspecific low back pain for more than 1 day; Netherlands; Occupational; January 1st 2009 - January 1st 2001
Sample size	615
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Expected duration of sick leave > 10 days (Yes/No)
Notes	Funding Source: Not reported Conflict of Interest: Not reported

Tran 2015

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People with chronic low back pain at an academic safety-net hospital, and affiliated community health centers; United States of America; Primary care; Data collection period not reported
Sample size	63 (twice-weekly yoga); 30 (once-weekly yoga)[a]
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: How helpful do you expect yoga to be for your back problems? (0 - 10)
Notes	Funding Source: The National Center for Complementary and Alternative Medicine; The National Institutes of Health Conflict of Interest: None declared [a]Tran 2015a (twice-weekly yoga) and Tran 2015b (once-weekly yoga) are same study, presented separately [b]Study does not present sufficient data to be included in meta-analyses (multivariate), but does provide some measure of statistical significance

Truchon 2012

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Workers receiving income replacement benefits because of common low back pain; Canada; Occupational; Data collection period not reported
Sample size	535
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Return to work expectations (time) (unclear scale)
Notes	Funding Source: Intitut de recherch Robert-Sauve en sante et en securite du travail Conflict of Interest: Not reported

Turner 2008

Population (low back pain description; location; setting; data collection period)	Prospective cohort: Workers with back injury claims with > 4 days of lost work time; United States of America; Occupational; July 2002 - April 2004
Sample size	1885
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Certainty of working in 6 months (0 - 10)
Notes	Funding Source: National Institute for Occupational Safety and Health, US Department of Health and Human Services Conflict of Interest: Not reported

Underwood 2007

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: Participants from general practices who consulted for simple low back pain that failed to resolve after their consultation; United Kingdom; Primary care; Data collection period not reported
Sample size	1334
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT: Treatment helpfulness (3 pts)
Notes	Funding Source: Not reported Conflict of Interest: None delcared

Van Hooff 2014

Population (low back pain description; location; setting; data collection period)	Prospective cohort: People with chronic low back pain, not improved with conservative care, referred to a tertiary orthopaedic spine care hospital; Netherlands; Setting not specified; October 2006 - January 2011
Sample size	524
Prognostic factor (type, measure(s) of individual recovery expectations)	SELF-EFFICACY: Pain Self-Efficacy Questionnaire (0 - 60)
Notes	Funding Source: Not reported Conflict of Interest: John O'Dowd is a direction of and shareholder in RealHealth Netherlands which was responsible for the assessments and treatment of patients [a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes [b]Study does not present sufficient data to be included in meta-analyses (multivariate), but does provide some measure of statistical significance

Van Wijk 2008

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trials (2 trials): People with back pain and sciatica recruited from pain clinics; Netherlands; Secondary care; Data collection period not reported
Sample size	81
Prognostic factor (type, measure(s) of individual recovery expectations)	GENERAL: Positive Expectations scale from Pain Cognitions List (5 pts)
Notes	Funding Source: Dutch Health Insurance Council; Pain Expertise Center Nijmegen Conflict of Interest: Not reported [a]Study does not present sufficient data to be included in meta-analyses (multivariate)

Yelland 2006

Population (low back pain description; location; setting; data collection period)	Randomized Controlled Trial: People with treatment-resistant chronic low back pain; Australia; Secondary care; Data collection period not reported
Sample size	110
Prognostic factor (type, measure(s) of individual recovery expectations)	TREATMENT (2): Desired improvement in function to make treatment worthwhile (0% - 100%), and Desired improvement in pain to make treatment worthwhile (0% - 100%)
Notes	Funding Source: Australian General Practice Evaluation Programme; The Australian Association of Musculoskeletal Medicine; The Musculoskeletal Research Foundation of Australia

Yelland 2006 *(Continued)*

Conflict of Interest: Not reported

[a]Study includes a recovery outcome based on categorisation of work, pain, and/or function outcomes

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Alyousef 2018	No or inappropriate measure of expectations
De Schepper 2016	No or inappropriate measure of expectations
Dozois 1995	No or inappropriate measure of expectations
Du Bois 2009	No appropriate follow-up period
Elfering 2015	No or inappropriate measure of expectations
Estlander 1998	Population not majority low back pain
Feuerstein 2006	No primary outcome
Fitzpatrick 1987	No or inappropriate measure of expectations
Goossens 2005	Population not majority low back pain
Hurwitz 2005	No or inappropriate measure of expectations
Jellema 2007	No primary outcome
Kendell 2018	No or inappropriate measure of expectations
Lochting 2017	No or inappropriate measure of expectations
Lotters 2006	Population not majority low back pain
Lurie 2016	Population not majority low back pain
Maxwell 1998	No appropriate follow-up period
Melloh 2011	No or inappropriate measure of expectations
Ng 2017	No or inappropriate measure of expectations
Reis 2007	No or inappropriate measure of expectations
Roberts 2015	No or inappropriate measure of expectations
Schultz 2008	No or inappropriate measure of expectations
Silvis 2016	No or inappropriate measure of expectations
Skargren 1998	Population not majority low back pain

Study	Reason for exclusion
Smeets 2008	Expectations not measured at baseline (after first treatment)
Soucy 2006	No or inappropriate measure of expectations
Staerkle 2004	No or inappropriate measure of expectations
Steenstra 2015	No or inappropriate measure of expectations
Vargas-Prada 2013	No or inappropriate measure of expectations
Wolff 2018	No or inappropriate measure of expectations

Characteristics of studies awaiting assessment *[ordered by study ID]*

Ailliet 2018

Notes	Likely to be eligible for inclusion

Ashworth 2013

Notes	Likely to be linked to an included study

Barons 2014

Notes	Likely to be eligible for inclusion

Friedman 2018

Notes	Likely to be eligible for inclusion

Ganesh 2019

Notes	Likely to be eligible for inclusion

Glattacker 2018

Notes	Likely to be eligible for inclusion

Harter 2004

Notes	German language

Hartvigsen 2018a

Notes	Likely to be linked to an included study

Klyne 2018

Notes	Likely to be eligible for inclusion

Mehling 2015

Notes	Likely to be eligible for inclusion

Mendelson 1983

Notes	Likely to be eligible for inclusion

Pfingsten 1997

Notes	Likely to be eligible for inclusion

Pfingsten 1997a

Notes	German language

Pfingsten 1997b

Notes	German language

Thomas 2005

Notes	Likely to be eligible for inclusion

Weber 1998

Notes	German language

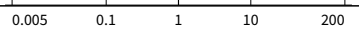
DATA AND ANALYSES

Comparison 1. Are expectations associated with work participation (closest to 12 months)?

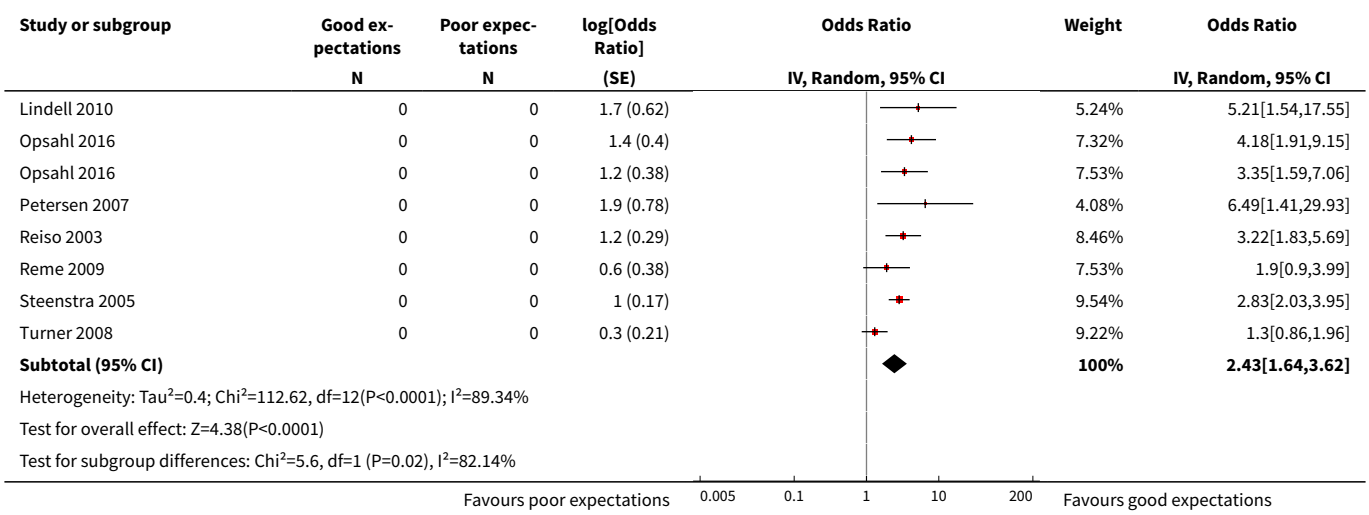
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Dichotomous measure of expectations	13		Odds Ratio (Random, 95% CI)	Subtotals only
1.1 Unadjusted results	10		Odds Ratio (Random, 95% CI)	4.11 [3.46, 4.89]
1.2 Adjusted results	12		Odds Ratio (Random, 95% CI)	2.43 [1.64, 3.62]
2 Continuous measure of expectations (/10)	2		Odds Ratio (Random, 95% CI)	Subtotals only
2.1 Unadjusted results	2		Odds Ratio (Random, 95% CI)	1.84 [0.53, 6.37]
2.2 Adjusted results	2		Odds Ratio (Random, 95% CI)	1.14 [0.95, 1.37]

Analysis 1.1. Comparison 1 Are expectations associated with work participation (closest to 12 months)?, Outcome 1 Dichotomous measure of expectations.

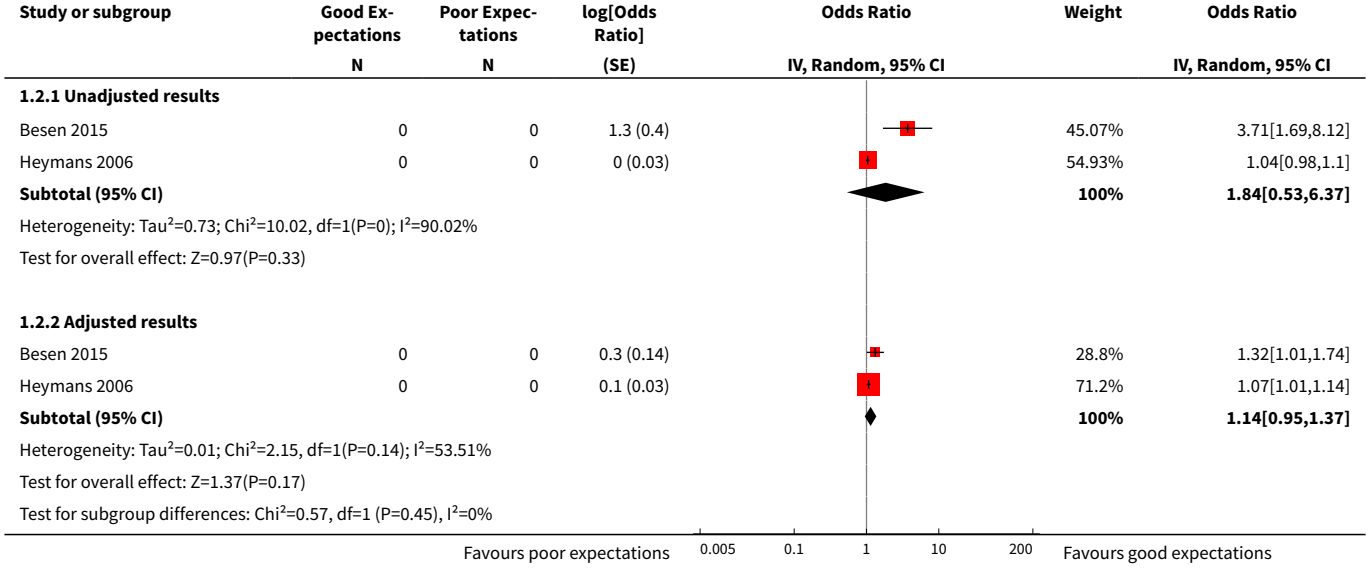
Study or subgroup	Good expectations N	Poor expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
1.1.1 Unadjusted results						
Carriere 2015	0	0	2.6 (0.79)		1.25%	13.6[2.89,63.97]
Du Bois 2008	0	0	1.7 (0.39)		4.99%	5.21[2.42,11.18]
Hagen 2005	0	0	0.8 (0.25)		11.6%	2.29[1.4,3.74]
Jensen 2013	0	0	1.2 (0.38)		5.25%	3.25[1.55,6.85]
Lindell 2010	0	0	1.9 (0.61)		2.08%	6.42[1.94,21.23]
Magnussen 2007	0	0	1.9 (0.66)		1.78%	6.89[1.89,25.12]
Opsahl 2016	0	0	1.6 (0.34)		6.5%	4.81[2.47,9.36]
Opsahl 2016	0	0	1.7 (0.33)		6.88%	5.37[2.81,10.24]
Reiso 2003	0	0	1.4 (0.56)		2.46%	3.97[1.33,11.91]
Steenstra 2005	0	0	1.4 (0.14)		31.7%	4.06[3.08,5.34]
Turner 2008	0	0	1.5 (0.16)		25.52%	4.31[3.15,5.89]
Subtotal (95% CI)					**100%**	**4.11[3.46,4.89]**
Heterogeneity: Tau²=0.01; Chi²=10.6, df=10(P=0.39); I²=5.62%						
Test for overall effect: Z=15.96(P<0.0001)						
1.1.2 Adjusted results						
Butler 2007	0	0	0.1 (0.41)		7.22%	1.15[0.52,2.57]
Carriere 2015	0	0	0 (0.01)		10.23%	1.04[1.02,1.06]
Du Bois 2008	0	0	1.5 (0.41)		7.22%	4.62[2.07,10.31]
Hagen 2005	0	0	0.6 (0.3)		8.36%	1.8[1,3.25]
Jensen 2013	0	0	1.1 (0.33)		8.05%	2.94[1.54,5.62]



Favours poor expectations — Favours good expectations



Analysis 1.2. Comparison 1 Are expectations associated with work participation (closest to 12 months)?, Outcome 2 Continuous measure of expectations (/10).



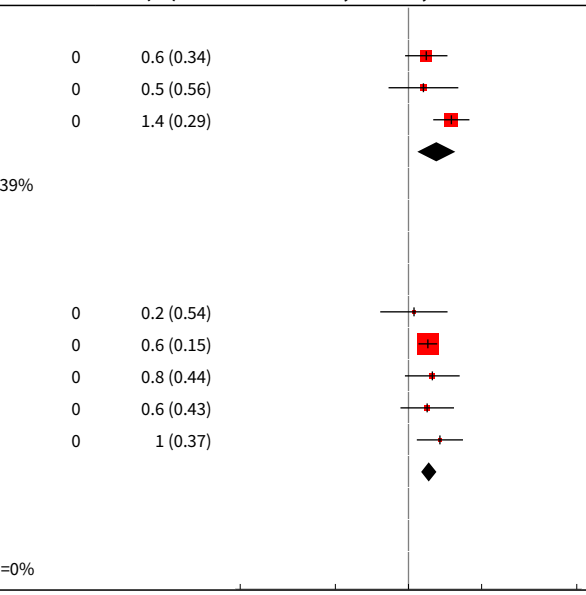
Comparison 2. Are expectations associated with important recovery (closest to 12 months)?

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Dichotomous measure of expectations	6		Odds Ratio (Random, 95% CI)	Subtotals only
1.1 Unadjusted results	3		Odds Ratio (Random, 95% CI)	2.40 [1.32, 4.37]
1.2 Adjusted results	5		Odds Ratio (Random, 95% CI)	1.89 [1.49, 2.41]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2 Continuous measure of expectations (/10)	4		Odds Ratio (Random, 95% CI)	Subtotals only
2.1 Unadjusted results	1		Odds Ratio (Random, 95% CI)	1.13 [1.11, 1.15]
2.2 Adjusted results	4		Odds Ratio (Random, 95% CI)	1.15 [1.07, 1.24]

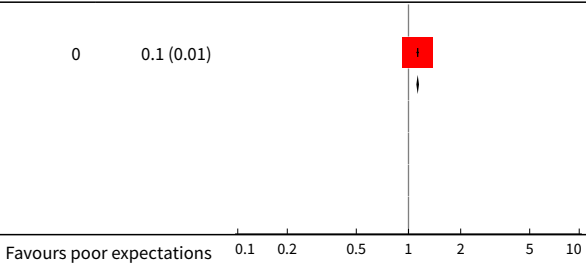
Analysis 2.1. Comparison 2 Are expectations associated with important recovery (closest to 12 months)?, Outcome 1 Dichotomous measure of expectations.

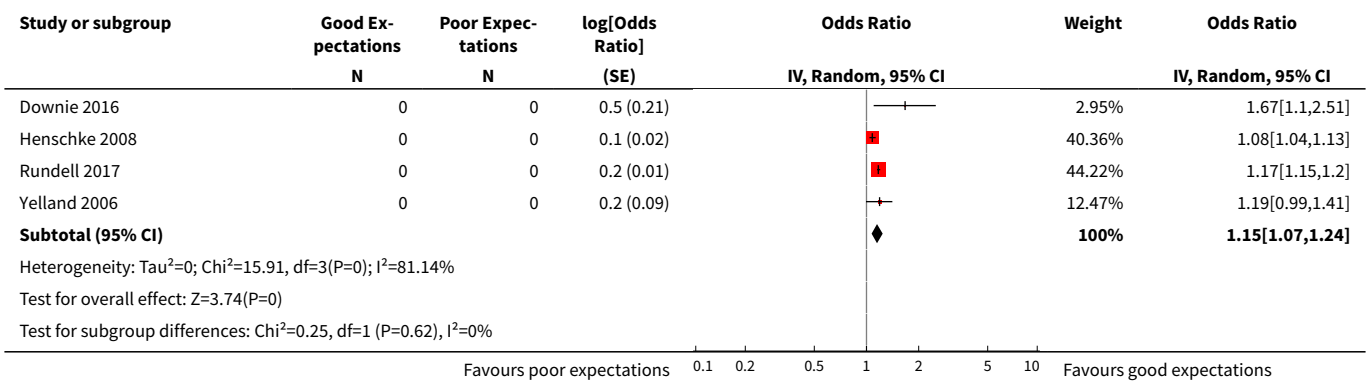
Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
2.1.1 Unadjusted results						
Niemisto 2004	0	0	0.6 (0.34)		36.99%	1.75[0.9,3.41]
Rasmussen-Barr 2012	0	0	0.5 (0.56)		20.76%	1.6[0.53,4.8]
Shaw 2009	0	0	1.4 (0.29)		42.25%	3.86[2.18,6.81]
Subtotal (95% CI)					**100%**	**2.4[1.32,4.37]**
Heterogeneity: Tau^2=0.14; Chi^2=3.95, df=2(P=0.14); I^2=49.39%						
Test for overall effect: Z=2.86(P=0)						
2.1.2 Adjusted results						
Enthoven 2006	0	0	0.2 (0.54)		5.22%	1.19[0.41,3.42]
Foster 2008	0	0	0.6 (0.15)		67.59%	1.84[1.37,2.47]
Niemisto 2004	0	0	0.8 (0.44)		7.86%	2.12[0.89,5.01]
Petersen 2007	0	0	0.6 (0.43)		8.23%	1.8[0.78,4.19]
Shaw 2009	0	0	1 (0.37)		11.11%	2.69[1.3,5.56]
Subtotal (95% CI)					**100%**	**1.89[1.49,2.41]**
Heterogeneity: Tau^2=0; Chi^2=1.77, df=4(P=0.78); I^2=0%						
Test for overall effect: Z=5.18(P<0.0001)						
Test for subgroup differences: Chi^2=0.51, df=1 (P=0.47), I^2=0%						



Analysis 2.2. Comparison 2 Are expectations associated with important recovery (closest to 12 months)?, Outcome 2 Continuous measure of expectations (/10).

Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
2.2.1 Unadjusted results						
Henschke 2008	0	0	0.1 (0.01)		100%	1.13[1.11,1.15]
Subtotal (95% CI)					**100%**	**1.13[1.11,1.15]**
Heterogeneity: Not applicable						
Test for overall effect: Z=12(P<0.0001)						
2.2.2 Adjusted results						

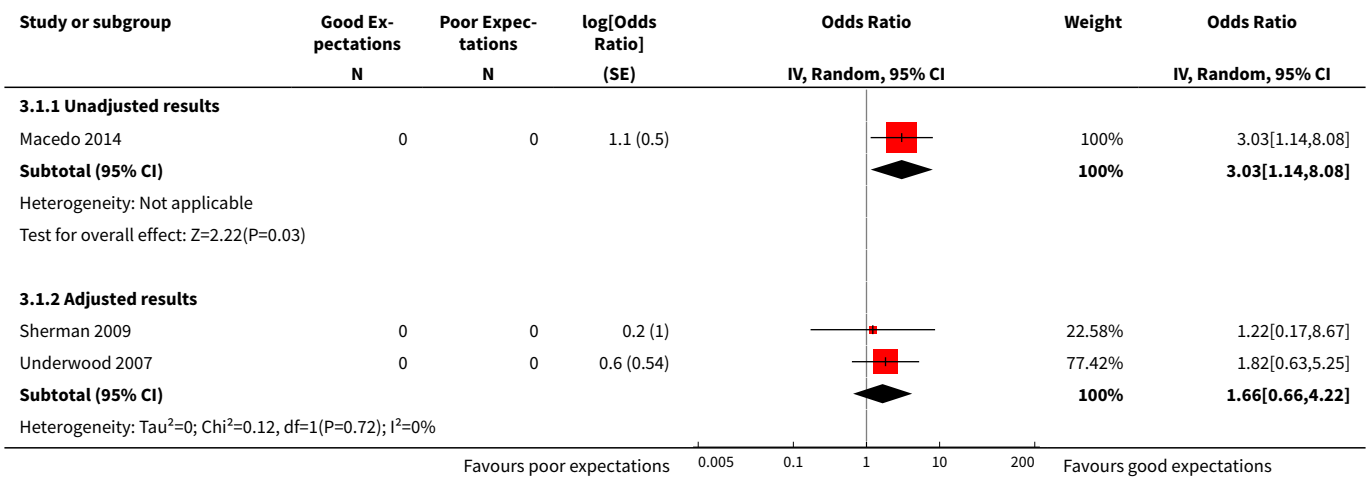


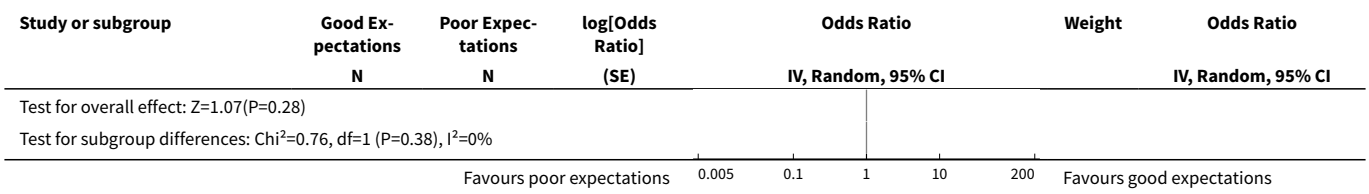


Comparison 3. Are expectations associated with functional limitation outcomes (closest to 12 months)?

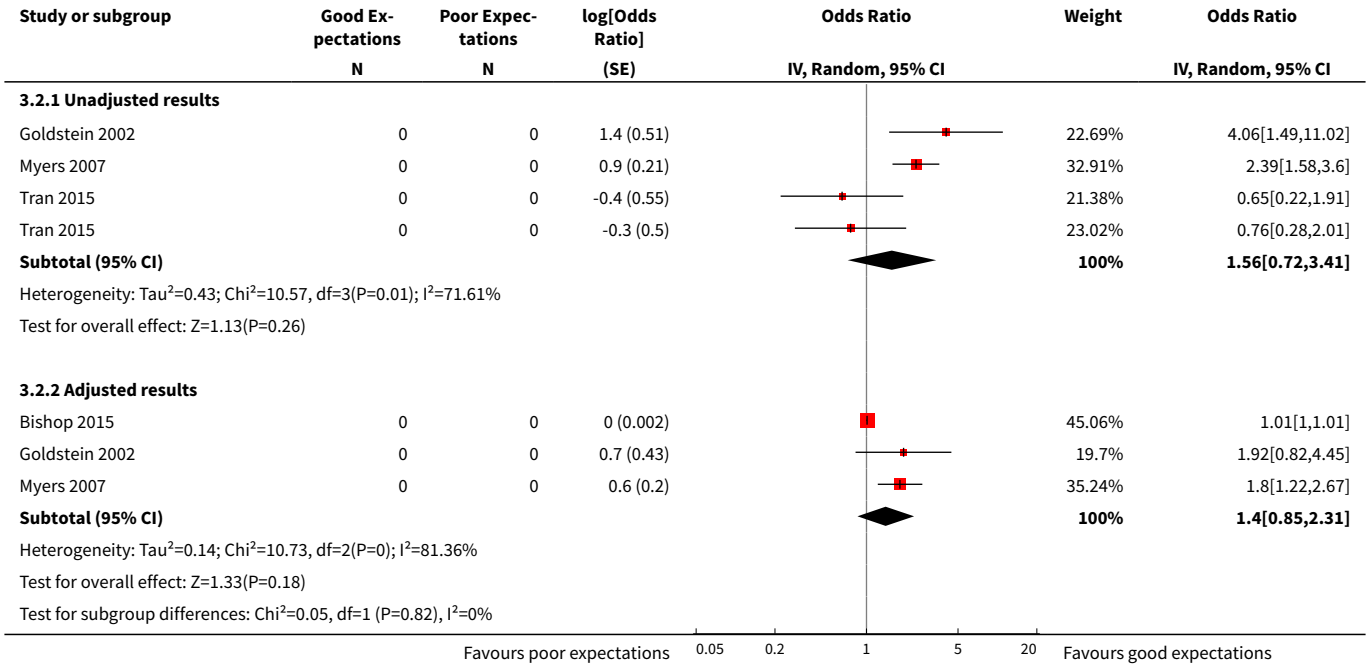
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Dichotomous measure of expectations	3		Odds Ratio (Random, 95% CI)	Subtotals only
1.1 Unadjusted results	1		Odds Ratio (Random, 95% CI)	3.03 [1.14, 8.08]
1.2 Adjusted results	2		Odds Ratio (Random, 95% CI)	1.66 [0.66, 4.22]
2 Continuous measure of expectations (/10)	4		Odds Ratio (Random, 95% CI)	Subtotals only
2.1 Unadjusted results	3		Odds Ratio (Random, 95% CI)	1.56 [0.72, 3.41]
2.2 Adjusted results	3		Odds Ratio (Random, 95% CI)	1.40 [0.85, 2.31]

Analysis 3.1. Comparison 3 Are expectations associated with functional limitation outcomes (closest to 12 months)?, Outcome 1 Dichotomous measure of expectations.





Analysis 3.2. Comparison 3 Are expectations associated with functional limitation outcomes (closest to 12 months)?, Outcome 2 Continuous measure of expectations (/10).



Comparison 4. Are expectations associated with pain outcomes (closest to 12 months)?

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Dichotomous measure of expectations	1		Odds Ratio (Random, 95% CI)	Subtotals only
1.1 Unadjusted results	1		Odds Ratio (Random, 95% CI)	2.51 [0.81, 7.82]
1.2 Adjusted results	0		Odds Ratio (Random, 95% CI)	0.0 [0.0, 0.0]
2 Continuous measure of expectations (/10)	3		Odds Ratio (Random, 95% CI)	Subtotals only
2.1 Unadjusted results	2		Odds Ratio (Random, 95% CI)	1.13 [0.48, 2.67]
2.2 Adjusted results	2		Odds Ratio (Random, 95% CI)	1.15 [1.08, 1.23]

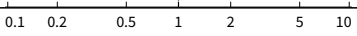
Analysis 4.1. Comparison 4 Are expectations associated with pain outcomes (closest to 12 months)?, Outcome 1 Dichotomous measure of expectations.

Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
4.1.1 Unadjusted results						
Macedo 2014	0	0	0.9 (0.58)		100%	2.51[0.81,7.82]
Subtotal (95% CI)					**100%**	**2.51[0.81,7.82]**
Heterogeneity: Tau²=0; Chi²=0, df=0(P<0.0001); I²=100%						
Test for overall effect: Z=1.59(P=0.11)						
4.1.2 Adjusted results						
Subtotal (95% CI)						**Not estimable**
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
Test for subgroup differences: Not applicable						

Favours poor expectations — Favours good expectations

Analysis 4.2. Comparison 4 Are expectations associated with pain outcomes (closest to 12 months)?, Outcome 2 Continuous measure of expectations (/10).

Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
4.2.1 Unadjusted results						
Goldstein 2002	0	0	0.8 (0.22)		39.58%	2.23[1.45,3.43]
Tran 2015	0	0	-0.4 (0.44)		30.42%	0.65[0.27,1.54]
Tran 2015	0	0	-0.2 (0.45)		30%	0.82[0.34,1.98]
Subtotal (95% CI)					**100%**	**1.13[0.48,2.67]**
Heterogeneity: Tau²=0.43; Chi²=8.61, df=2(P=0.01); I²=76.77%						
Test for overall effect: Z=0.29(P=0.77)						
4.2.2 Adjusted results						
Goldstein 2002	0	0	0.3 (0.19)		3.23%	1.32[0.91,1.92]
Kongsted 2014	0	0	0.2 (0.07)		23.82%	1.21[1.05,1.39]
Kongsted 2014	0	0	0.1 (0.04)		72.95%	1.13[1.04,1.22]
Subtotal (95% CI)					**100%**	**1.15[1.08,1.23]**
Heterogeneity: Tau²=0; Chi²=1.3, df=2(P=0.52); I²=0%						
Test for overall effect: Z=4.15(P<0.0001)						
Test for subgroup differences: Chi²=0, df=1 (P=0.97), I²=0%						



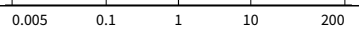
Favours poor expectations — Favours good expectations

Comparison 5. Work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted): Subgroup analyses

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Low back pain duration	12		Odds Ratio (Random, 95% CI)	Subtotals only
1.1 Acute low back pain population	0		Odds Ratio (Random, 95% CI)	0.0 [0.0, 0.0]
1.2 Subacute/Chronic low back pain population	4		Odds Ratio (Random, 95% CI)	2.34 [1.61, 3.40]
1.3 Mixed/Non-specified low back pain population	8		Odds Ratio (Random, 95% CI)	2.41 [1.48, 3.91]
2 Recovery expectation types	12		Odds Ratio (Random, 95% CI)	Subtotals only
2.1 General recovery expectations	12		Odds Ratio (Random, 95% CI)	2.43 [1.64, 3.62]
3 Recovery expectation reference time periods	12		Odds Ratio (Random, 95% CI)	Subtotals only
3.1 Short (1 month or less)	4		Odds Ratio (Random, 95% CI)	2.02 [1.00, 4.09]
3.2 Long (3-6 months)	4		Odds Ratio (Random, 95% CI)	2.83 [1.36, 5.89]
3.3 No / unclear reference period	4		Odds Ratio (Random, 95% CI)	2.55 [1.53, 4.25]
4 Outcome follow-up periods	12		Odds Ratio (Random, 95% CI)	Subtotals only
4.1 Short term follow-up (3 mo)	3		Odds Ratio (Random, 95% CI)	3.19 [1.77, 5.75]
4.2 Moderate term follow-up (5-8 mo; closest to 6 mo)	4		Odds Ratio (Random, 95% CI)	2.43 [1.58, 3.74]
4.3 Long term follow-up (8-16 mo; closest to 12 mo)	9		Odds Ratio (Random, 95% CI)	2.12 [1.41, 3.17]
4.4 Very long term follow-up (>16 mo)	2		Odds Ratio (Random, 95% CI)	3.10 [1.87, 5.12]
5 Study phase of investigation	12		Odds Ratio (Random, 95% CI)	Subtotals only
5.1 Exploratory analyses	10		Odds Ratio (Random, 95% CI)	2.41 [1.76, 3.29]
5.2 Confirmatory analyses	2		Odds Ratio (Random, 95% CI)	2.31 [0.82, 6.51]

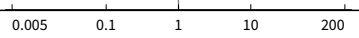
Analysis 5.1. Comparison 5 Work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted): Subgroup analyses, Outcome 1 Low back pain duration.

Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
5.1.1 Acute low back pain population						
Subtotal (95% CI)						**Not estimable**
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
5.1.2 Subacute/Chronic low back pain population						
Hagen 2005	0	0	0.6 (0.3)		36.48%	1.8[1,3.25]
Jensen 2013	0	0	1.1 (0.33)		30.67%	2.94[1.54,5.62]
Lindell 2010	0	0	1.7 (0.62)		9.24%	5.21[1.54,17.55]
Reme 2009	0	0	0.6 (0.38)		23.61%	1.9[0.9,3.99]
Subtotal (95% CI)					**100%**	**2.34[1.61,3.4]**
Heterogeneity: Tau^2=0.01; Chi^2=3.21, df=3(P=0.36); I^2=6.44%						
Test for overall effect: Z=4.45(P<0.0001)						
5.1.3 Mixed/Non-specified low back pain population						
Butler 2007	0	0	0.1 (0.41)		10.26%	1.15[0.52,2.57]
Carriere 2015	0	0	0 (0.01)		14.24%	1.04[1.02,1.06]
Du Bois 2008	0	0	1.5 (0.41)		10.26%	4.62[2.07,10.31]
Opsahl 2016	0	0	1.2 (0.38)		10.68%	3.35[1.59,7.06]
Opsahl 2016	0	0	1.4 (0.4)		10.4%	4.18[1.91,9.15]
Petersen 2007	0	0	1.9 (0.78)		5.92%	6.49[1.41,29.93]
Reiso 2003	0	0	1.2 (0.29)		11.93%	3.22[1.83,5.69]
Steenstra 2005	0	0	1 (0.17)		13.36%	2.83[2.03,3.95]
Turner 2008	0	0	0.3 (0.21)		12.93%	1.3[0.86,1.96]
Subtotal (95% CI)					**100%**	**2.41[1.48,3.91]**
Heterogeneity: Tau^2=0.43; Chi^2=90.54, df=8(P<0.0001); I^2=91.16%						
Test for overall effect: Z=3.54(P=0)						
Test for subgroup differences: Chi^2=0.01, df=1 (P=0.93), I^2=0%						



Analysis 5.2. Comparison 5 Work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted): Subgroup analyses, Outcome 2 Recovery expectation types.

Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
5.2.1 General recovery expectations						
Butler 2007	0	0	0.1 (0.41)		7.22%	1.15[0.52,2.57]
Carriere 2015	0	0	0 (0.01)		10.23%	1.04[1.02,1.06]
Du Bois 2008	0	0	1.5 (0.41)		7.22%	4.62[2.07,10.31]
Hagen 2005	0	0	0.6 (0.3)		8.36%	1.8[1,3.25]
Jensen 2013	0	0	1.1 (0.33)		8.05%	2.94[1.54,5.62]
Lindell 2010	0	0	1.7 (0.62)		5.24%	5.21[1.54,17.55]
Opsahl 2016	0	0	1.2 (0.38)		7.53%	3.35[1.59,7.06]
Opsahl 2016	0	0	1.4 (0.4)		7.32%	4.18[1.91,9.15]
Petersen 2007	0	0	1.9 (0.78)		4.08%	6.49[1.41,29.93]



Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
Reiso 2003	0	0	1.2 (0.29)		8.46%	3.22[1.83,5.69]
Reme 2009	0	0	0.6 (0.38)		7.53%	1.9[0.9,3.99]
Steenstra 2005	0	0	1 (0.17)		9.54%	2.83[2.03,3.95]
Turner 2008	0	0	0.3 (0.21)		9.22%	1.3[0.86,1.96]
Subtotal (95% CI)					**100%**	**2.43[1.64,3.62]**
Heterogeneity: Tau²=0.4; Chi²=112.62, df=12(P<0.0001); I²=89.34%						
Test for overall effect: Z=4.38(P<0.0001)						

Favours poor expectations 0.005 0.1 1 10 200 Favours good expectations

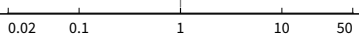
Analysis 5.3. Comparison 5 Work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted): Subgroup analyses, Outcome 3 Recovery expectation reference time periods.

Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
5.3.1 Short (1 month or less)						
Carriere 2015	0	0	0 (0.01)		28.18%	1.04[1.02,1.06]
Reiso 2003	0	0	1.2 (0.29)		23.84%	3.22[1.83,5.69]
Reme 2009	0	0	0.6 (0.38)		21.46%	1.9[0.9,3.99]
Steenstra 2005	0	0	1 (0.17)		26.52%	2.83[2.03,3.95]
Subtotal (95% CI)					**100%**	**2.02[1,4.09]**
Heterogeneity: Tau²=0.46; Chi²=52.01, df=3(P<0.0001); I²=94.23%						
Test for overall effect: Z=1.95(P=0.05)						
5.3.2 Long (3-6 months)						
Du Bois 2008	0	0	1.5 (0.41)		25.23%	4.62[2.07,10.31]
Jensen 2013	0	0	1.1 (0.33)		28.23%	2.94[1.54,5.62]
Petersen 2007	0	0	1.9 (0.78)		14.09%	6.49[1.41,29.93]
Turner 2008	0	0	0.3 (0.21)		32.46%	1.3[0.86,1.96]
Subtotal (95% CI)					**100%**	**2.83[1.36,5.89]**
Heterogeneity: Tau²=0.39; Chi²=12.06, df=3(P=0.01); I²=75.13%						
Test for overall effect: Z=2.77(P=0.01)						
5.3.3 No / unclear reference period						
Butler 2007	0	0	0.1 (0.41)		19.97%	1.15[0.52,2.57]
Hagen 2005	0	0	0.6 (0.3)		25.88%	1.8[1,3.25]
Lindell 2010	0	0	1.7 (0.62)		12.24%	5.21[1.54,17.55]
Opsahl 2016	0	0	1.4 (0.4)		20.46%	4.18[1.91,9.15]
Opsahl 2016	0	0	1.2 (0.38)		21.46%	3.35[1.59,7.06]
Subtotal (95% CI)					**100%**	**2.55[1.53,4.25]**
Heterogeneity: Tau²=0.17; Chi²=8.38, df=4(P=0.08); I²=52.26%						
Test for overall effect: Z=3.58(P=0)						
Test for subgroup differences: Chi²=0.45, df=1 (P=0.8), I²=0%						

Favours poor expectations 0.005 0.1 1 10 200 Favours good expectations

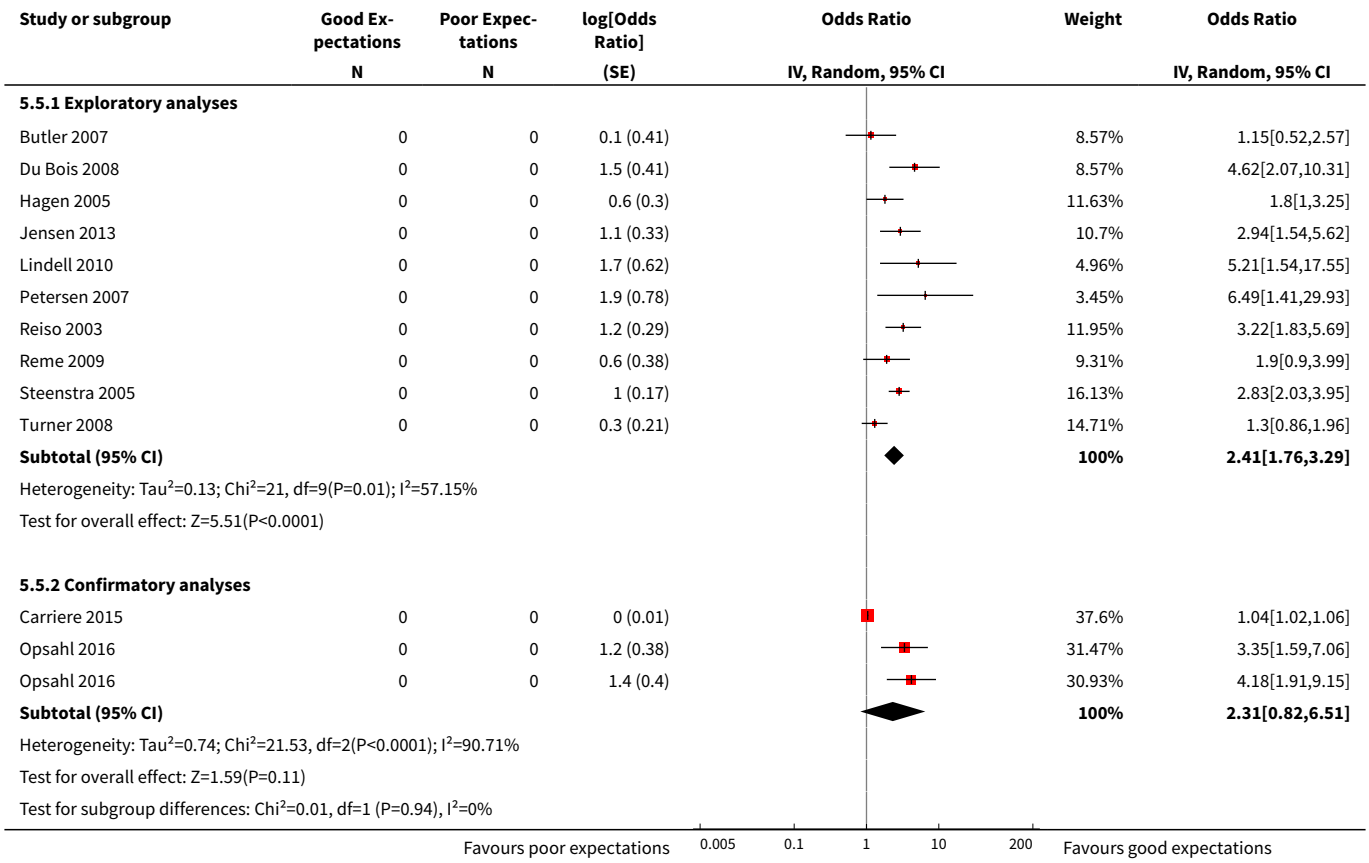
Analysis 5.4. Comparison 5 Work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted): Subgroup analyses, Outcome 4 Outcome follow-up periods.

Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
5.4.1 Short term follow-up (3 mo)						
Du Bois 2008	0	0	1.5 (0.41)		32.41%	4.62[2.07,10.31]
Hagen 2005	0	0	0.6 (0.35)		38.72%	1.9[0.96,3.77]
Reme 2009	0	0	1.4 (0.45)		28.86%	4.22[1.75,10.2]
Subtotal (95% CI)					**100%**	**3.19[1.77,5.75]**
Heterogeneity: Tau²=0.11; Chi²=3.39, df=2(P=0.18); I²=40.93%						
Test for overall effect: Z=3.85(P=0)						
5.4.2 Moderate term follow-up (5-8 mo; closest to 6 mo)						
Butler 2007	0	0	0.3 (0.31)		26.92%	1.34[0.73,2.45]
Lindell 2010	0	0	1.4 (0.68)		8.86%	4.1[1.08,15.53]
Steenstra 2005	0	0	1 (0.17)		43%	2.83[2.03,3.95]
Turner 2008	0	0	1.1 (0.38)		21.22%	3.06[1.46,6.45]
Subtotal (95% CI)					**100%**	**2.43[1.58,3.74]**
Heterogeneity: Tau²=0.08; Chi²=5.42, df=3(P=0.14); I²=44.7%						
Test for overall effect: Z=4.04(P<0.0001)						
5.4.3 Long term follow-up (8-16 mo; closest to 12 mo)						
Butler 2007	0	0	0.1 (0.41)		9.4%	1.15[0.52,2.57]
Carriere 2015	0	0	0 (0.01)		14.95%	1.04[1.02,1.06]
Hagen 2005	0	0	0.6 (0.3)		11.36%	1.8[1,3.25]
Jensen 2013	0	0	1.1 (0.33)		10.81%	2.94[1.54,5.62]
Lindell 2010	0	0	1.7 (0.62)		6.36%	5.21[1.54,17.55]
Opsahl 2016	0	0	1.4 (0.4)		9.57%	4.18[1.91,9.15]
Opsahl 2016	0	0	1.2 (0.38)		9.92%	3.35[1.59,7.06]
Petersen 2007	0	0	1.9 (0.78)		4.76%	6.49[1.41,29.93]
Reme 2009	0	0	0.6 (0.38)		9.92%	1.9[0.9,3.99]
Turner 2008	0	0	0.3 (0.21)		12.95%	1.3[0.86,1.96]
Subtotal (95% CI)					**100%**	**2.12[1.41,3.17]**
Heterogeneity: Tau²=0.28; Chi²=50.5, df=9(P<0.0001); I²=82.18%						
Test for overall effect: Z=3.64(P=0)						
5.4.4 Very long term follow-up (>16 mo)						
Lindell 2010	0	0	1 (0.55)		21.75%	2.69[0.92,7.91]
Reiso 2003	0	0	1.2 (0.29)		78.25%	3.22[1.83,5.69]
Subtotal (95% CI)					**100%**	**3.1[1.87,5.12]**
Heterogeneity: Tau²=0; Chi²=0.08, df=1(P=0.77); I²=0%						
Test for overall effect: Z=4.41(P<0.0001)						
Test for subgroup differences: Chi²=1.98, df=1 (P=0.58), I²=0%						



Favours poor expectations — Favours good expectations

Analysis 5.5. Comparison 5 Work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted): Subgroup analyses, Outcome 5 Study phase of investigation.



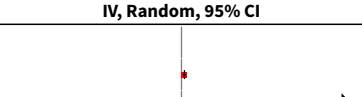
Comparison 6. QUIPS ROB domain ratings (work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted))

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 'Participation' domain	12		Odds Ratio (Random, 95% CI)	Subtotals only
1.1 Studies with low ROB	6		Odds Ratio (Random, 95% CI)	2.76 [1.48, 5.17]
1.2 Studies with moderate/high ROB	6		Odds Ratio (Random, 95% CI)	2.16 [1.41, 3.31]
2 'Attrition' domain	12		Odds Ratio (Random, 95% CI)	Subtotals only
2.1 Studies with low ROB	10		Odds Ratio (Random, 95% CI)	2.72 [2.05, 3.60]
2.2 Studies with moderate/high ROB	2		Odds Ratio (Random, 95% CI)	1.04 [1.02, 1.06]
3 'Prognostic factor' domain	12		Odds Ratio (Random, 95% CI)	Subtotals only

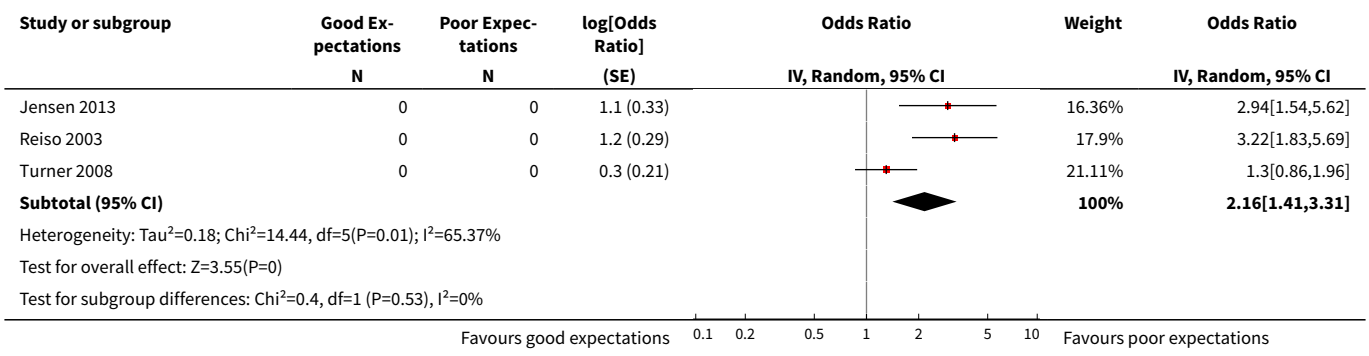
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 Studies with low ROB	0		Odds Ratio (Random, 95% CI)	0.0 [0.0, 0.0]
3.2 Studies with moderate/high ROB	12		Odds Ratio (Random, 95% CI)	2.43 [1.64, 3.62]
4 'Outcome' domain	12		Odds Ratio (Random, 95% CI)	Subtotals only
4.1 Studies with low ROB	10		Odds Ratio (Random, 95% CI)	2.41 [1.76, 3.29]
4.2 Studies with moderate/high ROB	2		Odds Ratio (Random, 95% CI)	2.31 [0.82, 6.51]
5 'Confounding' domain	12		Odds Ratio (Random, 95% CI)	Subtotals only
5.1 Studies with low ROB	8		Odds Ratio (Random, 95% CI)	2.03 [1.34, 3.07]
5.2 Studies with moderate/high ROB	4		Odds Ratio (Random, 95% CI)	3.10 [2.36, 4.07]
6 'Analysis and Reporting' domain	12		Odds Ratio (Random, 95% CI)	Subtotals only
6.1 Studies with low ROB	9		Odds Ratio (Random, 95% CI)	2.26 [1.46, 3.50]
6.2 Studies with moderate/high ROB	3		Odds Ratio (Random, 95% CI)	3.03 [1.91, 4.78]

Analysis 6.1. Comparison 6 QUIPS ROB domain ratings (work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted)), Outcome 1 'Participation' domain.

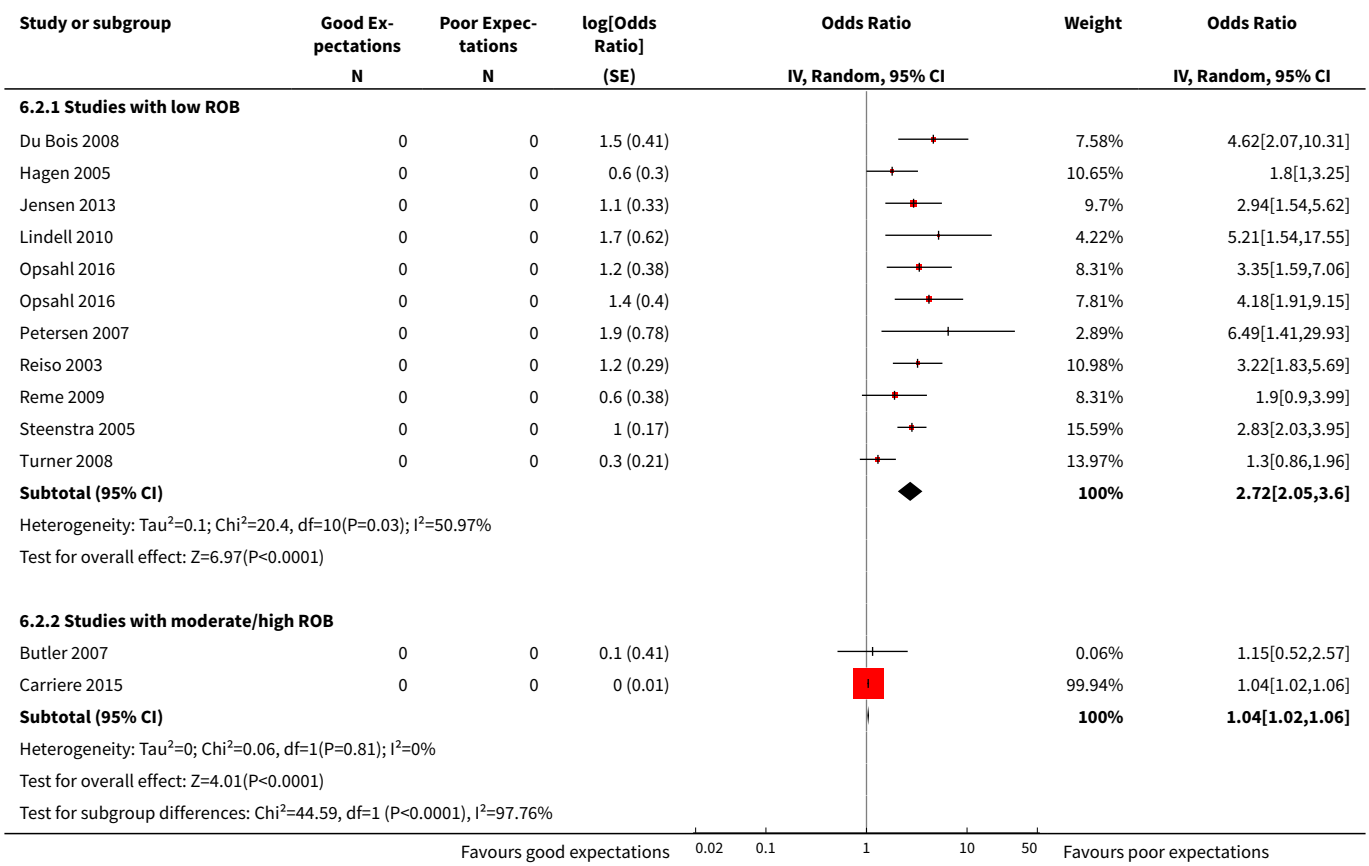
Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
6.1.1 Studies with low ROB						
Carriere 2015	0	0	0 (0.01)		18.62%	1.04[1.02,1.06]
Lindell 2010	0	0	1.7 (0.62)		10.95%	5.21[1.54,17.55]
Opsahl 2016	0	0	1.2 (0.38)		14.74%	3.35[1.59,7.06]
Opsahl 2016	0	0	1.4 (0.4)		14.42%	4.18[1.91,9.15]
Petersen 2007	0	0	1.9 (0.78)		8.83%	6.49[1.41,29.93]
Reme 2009	0	0	0.6 (0.38)		14.74%	1.9[0.9,3.99]
Steenstra 2005	0	0	1 (0.17)		17.69%	2.83[2.03,3.95]
Subtotal (95% CI)					**100%**	**2.76[1.48,5.17]**
Heterogeneity: Tau²=0.55; Chi²=70.51, df=6(P<0.0001); I²=91.49%						
Test for overall effect: Z=3.18(P=0)						
6.1.2 Studies with moderate/high ROB						
Butler 2007	0	0	0.1 (0.41)		13.57%	1.15[0.52,2.57]
Du Bois 2008	0	0	1.5 (0.41)		13.57%	4.62[2.07,10.31]
Hagen 2005	0	0	0.6 (0.3)		17.51%	1.8[1,3.25]



Favours good expectations — Favours poor expectations

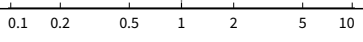


Analysis 6.2. Comparison 6 QUIPS ROB domain ratings (work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted)), Outcome 2 'Attrition' domain.



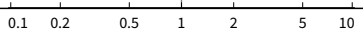
Analysis 6.3. Comparison 6 QUIPS ROB domain ratings (work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted)), Outcome 3 'Prognostic factor' domain.

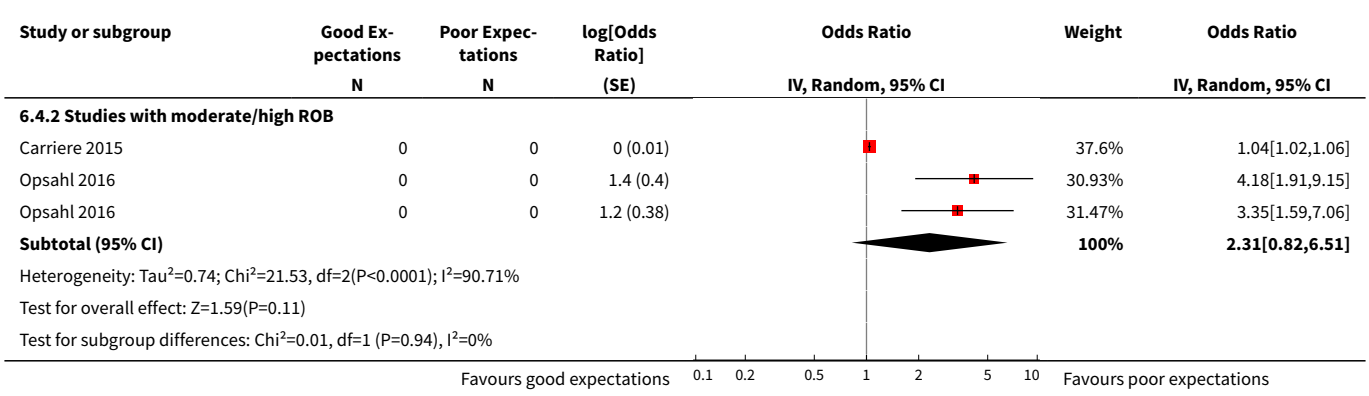
Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
6.3.1 Studies with low ROB						
Subtotal (95% CI)						**Not estimable**
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
6.3.2 Studies with moderate/high ROB						
Butler 2007	0	0	0.1 (0.41)		7.22%	1.15[0.52,2.57]
Carriere 2015	0	0	0 (0.01)		10.23%	1.04[1.02,1.06]
Du Bois 2008	0	0	1.5 (0.41)		7.22%	4.62[2.07,10.31]
Hagen 2005	0	0	0.6 (0.3)		8.36%	1.8[1,3.25]
Jensen 2013	0	0	1.1 (0.33)		8.05%	2.94[1.54,5.62]
Lindell 2010	0	0	1.7 (0.62)		5.24%	5.21[1.54,17.55]
Opsahl 2016	0	0	1.4 (0.4)		7.32%	4.18[1.91,9.15]
Opsahl 2016	0	0	1.2 (0.38)		7.53%	3.35[1.59,7.06]
Petersen 2007	0	0	1.9 (0.78)		4.08%	6.49[1.41,29.93]
Reiso 2003	0	0	1.2 (0.29)		8.46%	3.22[1.83,5.69]
Reme 2009	0	0	0.6 (0.38)		7.53%	1.9[0.9,3.99]
Steenstra 2005	0	0	1 (0.17)		9.54%	2.83[2.03,3.95]
Turner 2008	0	0	0.3 (0.21)		9.22%	1.3[0.86,1.96]
Subtotal (95% CI)					**100%**	**2.43[1.64,3.62]**
Heterogeneity: Tau^2=0.4; Chi^2=112.62, df=12(P<0.0001); I^2=89.34%						
Test for overall effect: Z=4.38(P<0.0001)						
Test for subgroup differences: Not applicable						



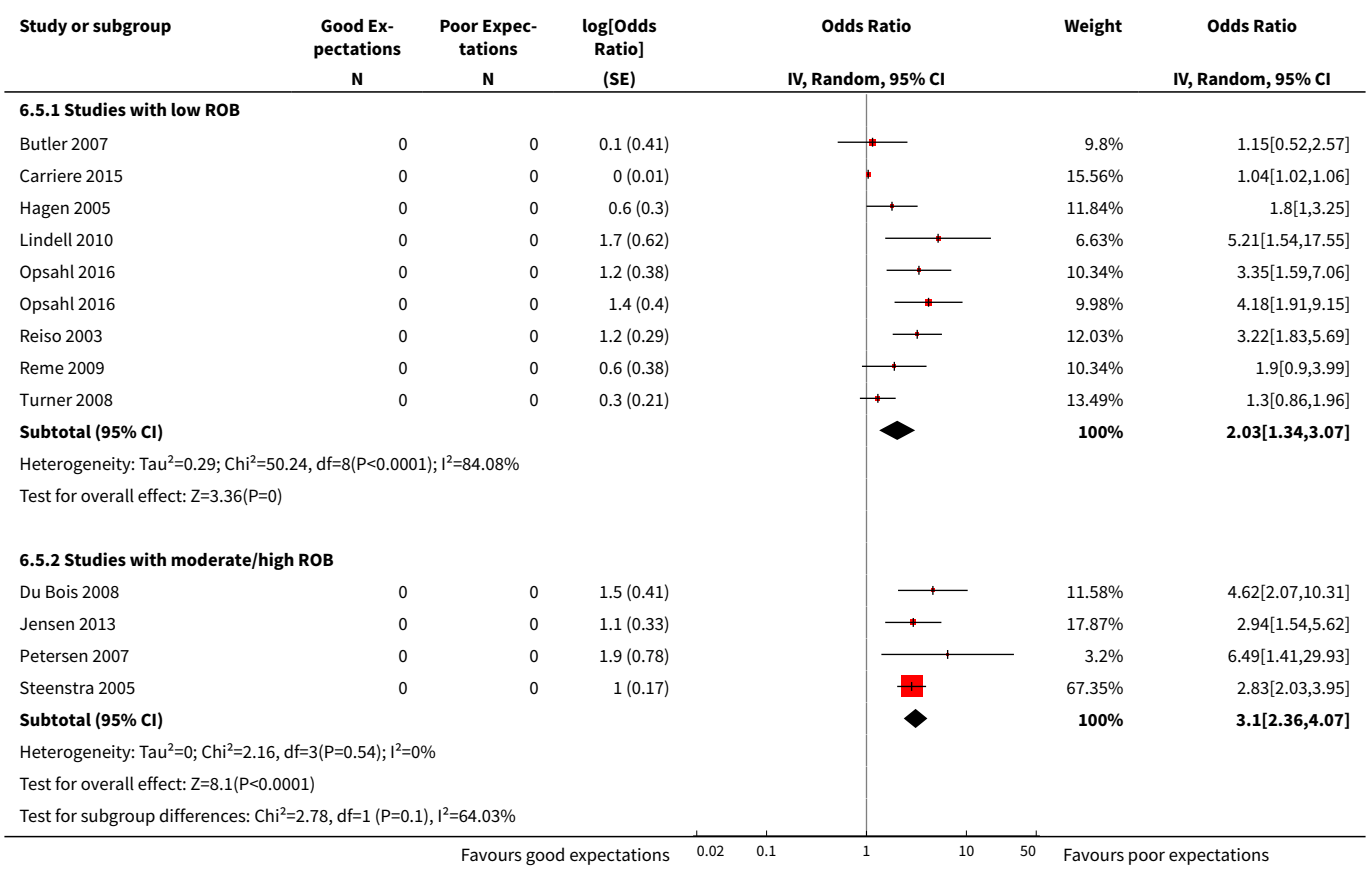
Analysis 6.4. Comparison 6 QUIPS ROB domain ratings (work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted)), Outcome 4 'Outcome' domain.

Study or subgroup	Good Expectations N	Poor Expectations N	log[Odds Ratio] (SE)	Odds Ratio IV, Random, 95% CI	Weight	Odds Ratio IV, Random, 95% CI
6.4.1 Studies with low ROB						
Butler 2007	0	0	0.1 (0.41)		8.57%	1.15[0.52,2.57]
Du Bois 2008	0	0	1.5 (0.41)		8.57%	4.62[2.07,10.31]
Hagen 2005	0	0	0.6 (0.3)		11.63%	1.8[1,3.25]
Jensen 2013	0	0	1.1 (0.33)		10.7%	2.94[1.54,5.62]
Lindell 2010	0	0	1.7 (0.62)		4.96%	5.21[1.54,17.55]
Petersen 2007	0	0	1.9 (0.78)		3.45%	6.49[1.41,29.93]
Reiso 2003	0	0	1.2 (0.29)		11.95%	3.22[1.83,5.69]
Reme 2009	0	0	0.6 (0.38)		9.31%	1.9[0.9,3.99]
Steenstra 2005	0	0	1 (0.17)		16.13%	2.83[2.03,3.95]
Turner 2008	0	0	0.3 (0.21)		14.71%	1.3[0.86,1.96]
Subtotal (95% CI)					**100%**	**2.41[1.76,3.29]**
Heterogeneity: Tau^2=0.13; Chi^2=21, df=9(P=0.01); I^2=57.15%						
Test for overall effect: Z=5.51(P<0.0001)						

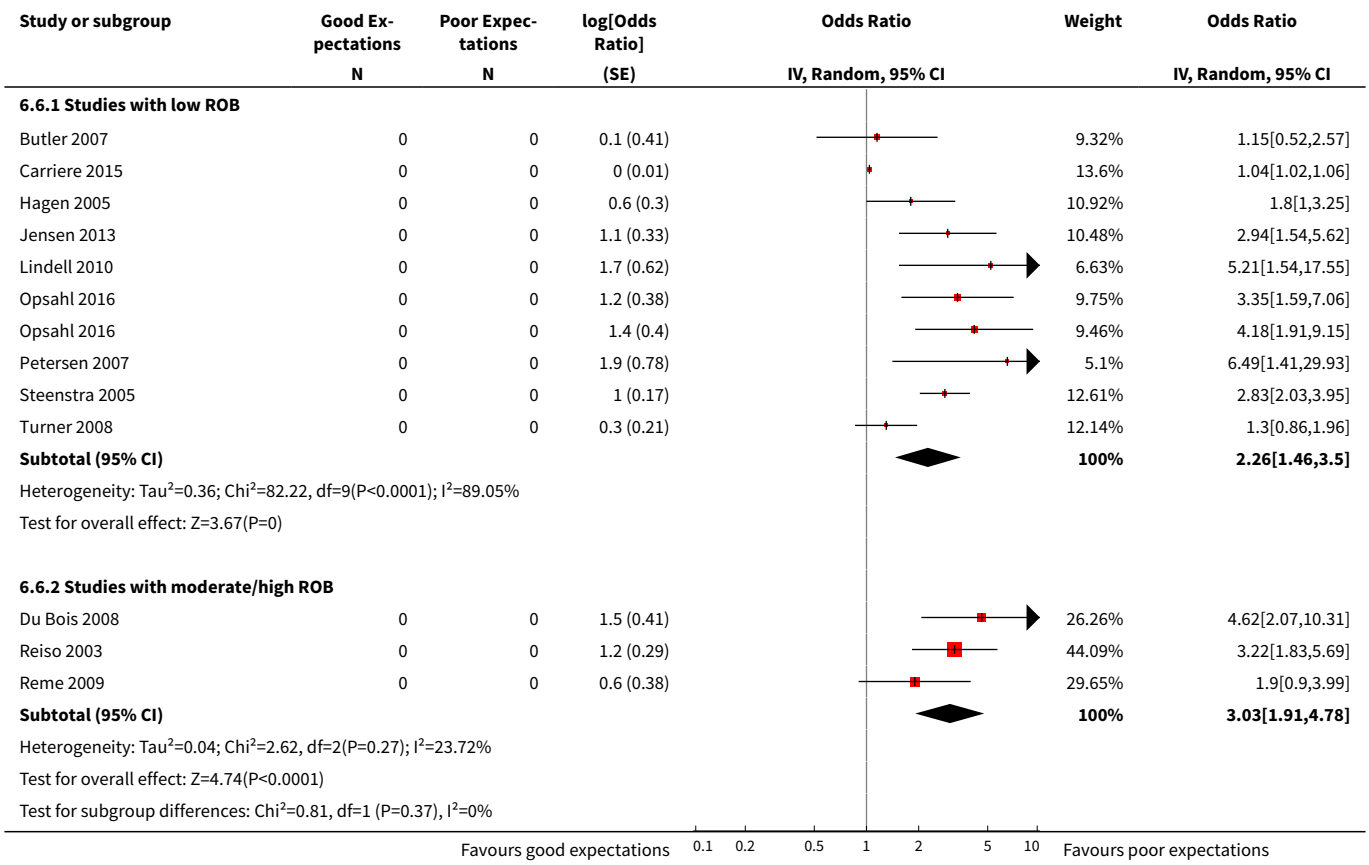




Analysis 6.5. Comparison 6 QUIPS ROB domain ratings (work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted)), Outcome 5 'Confounding' domain.



Analysis 6.6. Comparison 6 QUIPS ROB domain ratings (work participation outcome (closest to 12 months); dichotomous measure of expectations (adjusted)), Outcome 6 'Analysis and Reporting' domain.



ADDITIONAL TABLES

Table 1. Components of the systematic review question

Review question component	Description
Population	Adults with non-specific low back pain (grouped as acute (< 6 weeks), subacute/chronic (≥ 6 weeks), and mixed duration), in any setting (grouped as worker, healthcare and general population settings)
Prognostic factor/comparator	Individual recovery expectations, measured at an early point in management, with any reference period (grouped as 1 month, 6 months, or none/unclear reference period)
Primary outcomes	Work participation, functional limitations, important recovery, or pain intensity
Time periods	Short (closest to 3 months), medium (closest to 6 months), long follow-up (closest to 12 months)[a], and very long follow-up (greater than 16 months)

[a]12 month follow-up period prioritised for primary analyses

Table 2. Descriptive summary of included studies

Study variables		All included studies (n = 60)		Synthesis studies (n = 52)	
		n	%	n	%
Year published	Older (before 2013)	44	73.3%	37	71.2%
	Recent (2013 - 2018)	16	26.7%	16	30.8%
Sample size (median, IQR)		257	132 - 592	312	166 - 627
Population source	Healthcare	37	61.7%	33	63.5%
	Occupational	16	26.2%	14	26.4%
	General	3	4.9%	3	5.7%
	Mixed	3	4.9%	1	1.9%
	Not specified	1	1.6%	1	1.9%
Duration of LBP	Acute	5	8.2%	4	7.5%
	Subacute	5	8.2%	5	9.4%
	Chronic	22	36.7%	20	38.5%
	Mixed	19	31.1%	14	26.4%
	Not specified	9	14.8%	9	17.0%
Type of expectations measure	General expectations	42	68.9%	36	67.9%
	Self-efficacy expectations	12	19.7%	12	22.6%
	Treatment expectations	17	28.3%	14	26.9%
Number of expectations measures	1	44	73.3%	39	75.0%
	2	13	21.3%	11	20.8%
	3	2	3.3%	2	3.8%
	4	1	1.6%	0	0.0%
Prognostic factor study phase	Exploratory/TEM	3	5.0%	3	5.8%
	Exploratory	44	73.3%	37	71.2%
	Confirmatory	13	21.3%	12	22.6%
Outcomes assessed	Pain	31	51.7%	24	46.2%
	Functional limitations	36	60.0%	30	57.7%

Table 2. Descriptive summary of included studies *(Continued)*

	Work participation	35	58.3%	31	59.6%
	Satisfaction	5	8.3%	2	3.8%
	Global improvement	9	15.0%	8	15.4%
	Health-related quality of life	6	10.0%	4	7.7%
	Cost	7	11.5%	6	11.3%
	Mood	3	4.9%	3	5.7%
Follow-up times available	Short (3 - 4 months)	29	48.3%	24	46.2%
	Moderate (5 - 8 closest to 6 months)	16	26.2%	13	24.5%
	Long (8 - 16 closest to 12 months)	40	65.6%	35	66.0%
	Very long (> 16 months)	7	11.5%	7	13.2%
Low risk of bias by QUIPS domain	Study Participation	18	30.0%	17	32.7%
	Study Attrition	36	60.0%	33	63.5%
	Prognostic Factor Measurement	27	45.0%	24	46.2%
	Outcome Measurement	54	90.0%	46	88.5%
	Study Confounding	28	46.7%	27	51.9%
	Statistical Analyses & Reporting	41	68.3%	38	73.1%
All QUIPS ROB domains rated low or moderate		36	60.0%	34	65.4%
All QUIPS ROB domains rated low		2	3.3%	2	3.8%

LBP = low back pain; ROB = risk of bias; TEM = treatment effect modification; QUIPS = Quality in Prognosis Studies Tool

Table 3. Reported associations for studies measuring work participation outcomes

Study	Sample size	Expectation measure	Outcome measure	Outcome follow-up period	Study reported effect; variance measure	Common effect size (lnOR)[a]	Standard error	Reported direction of association[b]
Besen 2015	241	General: How soon do you expect to be able to resume your normal job without any limitations? (6-pt; 0 - 2 days up to > 60 days); reverse coded for analysis (higher = better)	Return to work	3 months	StB; P	1.24	0.63	+
Butler 2007	173	General: Expectations of recovery (5-pt), dichotomised as positive (≥ get better soon) vs negative (≤ get better slowly)	Unstable employment pattern	12 months	OR; P	0.14	0.41	Ø
Carriere 2015	108	General: Likelihood of return to work in next month (0 - 100), dichotomised as low (< 62.5) vs high (≥ 62.5)	Successful return to work	12 months	OR; 95% CI	0.04	0.01	+
Demmelmaier 2010	77	General: Pain expectations (sum of 2 7-pt rating scales; 1 adapted from OMPQ; 0 - 12, higher = worse)	On sick leave	12 months	P[c]	N/A[d]	N/A	Ø
Du Bois 2008	186	General: Return to work certainty within 6 months (0 - 10), dichotomised as not very sure (< 10) vs very sure (10)	Non-return to work	3 months	OR; 95% CI	1.53	0.41	+
Gervais 1991	135	Self-efficacy: Self-Efficacy and Results Expectancies Inventory (unclear scale; higher = better))	Non-full-time return to work	6 months	OR; P	−0.34	0.20	Ø
Gross 2010	298	General: Work-related Recovery Expectations Questionnaire (average of 3 Likert scales, 1 - 5, higher = worse)	Time to suspension of time-loss benefits	12 months	HR; 95% CI	0.19	0.07	+
Haldorsen 1998	84	General: Do you believe that you will be back to work after a couple of weeks? (5-pt scale, higher = worse)	Non-return to work	12 months	F, P	0.84[d]	0.43	+
Hagen 2005	457	General: Belief that back pain will disappear from Graded Reduced Work Ability Scale (1-6), dichotomized at median as don't believe back pain will disappear	Non-return to work	12 months	OR; 95% CI	0.59	0.30	+
Harkapaa 1996	175	General: Health Optimism Index (5-20); reverse coded for analyses (higher = better)	Return to work	12 months	P[c]	N/A	N/A	+

Table 3. Reported associations for studies measuring work participation outcomes *(Continued)*

Heymans 2006	268	General: When do you think you will be able to work full-time again? (8-pt, < 1 week to > 12 months, and no idea)	Time to full return to work	12 months	HR; 95% CI	0.05	0.02	+
Jensen 2000	107	Self-efficacy: Belief in ability for learning to cope with the pain (unclear scale, higher = better)	Receiving disability pension	6 months	OR; 95% CI	0.21	0.77	Ø
Jensen 2013	282	General: Expectations of return to work within 6 months (10-pt), dichotomised as not convinced (< 8) vs convinced (8 - 10)	Non-return to work	12 months	OR; 95% CI	1.08	0.33	+
Lindell 2010	123	General: Self-prediction of probability of return to work at some time in the future (5-pt Likert), dichotomised as high probability (≥ rather probable) vs low probability (≤ rather improbable)	Stable return to work	12 months	OR; 95% CI	1.65	0.62	+
Magnussen 2007	79	General: Do you believe that you will ever be able to return to work?, dichotomised as yes vs no/don't know	Entered a return to work process	12 months	OR; 95% CI	1.93[d]	0.66	+
Opsahl 2016**a**[e]	286	General: Predicted extent of return to work (4-pt), dichotomised as high vs low/moderate degree	Return to work	12 months	OR; 95% CI	1.21	0.38	+
Opsahl 2016**b**[e]	283	General: Predicted extent of return to work (4-pt), dichotomised as high vs low/moderate degree	Return to work	12 months	OR; 95% CI	1.43	0.40	+
Opsommer 2017	98	General: In your estimation, what are the chances that you will be able to work in 6 months? (0 - 10, higher = better)	Time to return to work	12 months	Harrell's C statistic; 95% CI	N/A[d]	N/A	+
Petersen 2007	153	General: Certainty of working 6 months after treatment (0 - 10), dichotomised at median as low vs high	Sick-listed	14 months	OR; 95% CI	1.87	0.78	+
Reiso 2003	153	General: Self-predicted work status in 4 weeks (3-pt), dichotomised as full return to work vs not full return to work	Return to work for at least 60 calendar days	24 months	HR; 95% CI	1.17	0.29	+
Reme 2009	173	General: Expectation to return to work within the next few weeks, dichotomised as negative (no/no opinion) vs positive (yes)	Non-return to work	12 months	OR; 95% CI	0.64	0.38	Ø

Table 3. Reported associations for studies measuring work participation outcomes *(Continued)*

Sandstrom 1986	52	Self-efficacy: I am afraid to start working again because I don't think I will be able to manage (7-pt)	Not sick-listed	12 months	MD; P	2.35[d]	0.71	+
Schultz 2004	214	General: Expectations of recovery scale (8 items, unclear scale, higher = worse)	Return to work	3 months	B; SE	0.25	0.06	+
Steenstra 2005	515	General: Expected duration of sick leave >10 days vs ≤ 10 days	Any return to work	6 months	HR; 95% CI	1.04	0.17	+
Truchon 2012	530	General: Return to work expectations (time, unclear scale, higher = worse)	Work absence	12 months	B; SE	0.25	0.00	+
Turner 2008	1885	General: Certainty of working in 6 months (0 - 10), dichotomised as low/no response (0 - 6) vs very high	Receiving wage replacement compensation	12 months	OR; 95% CI	0.26	0.21	Ø

Table 3. Description of the reported associations between the primary expectations measure and return to work participation outcomes, including presentation as common natural log odds effect size and standard error. Results presented are from the best adjusted multivariate model, when available, selecting the available study time period in study closest to 12 months (positive association in 19 studies (20 groups), no association in 6 studies).

[a]All reported associations have been converted to the natural log odds (lnOR) scale and the same direction when possible; lnOR > 1 indicates a positive direction of association between expectations and outcome.

[b]Direction of association: + = positive, associated with better outcome; Ø = neutral, no association with outcome; - = negative, associated with worse outcome

[c]Study where results are from unadjusted models.

[d]Statistical significance only reported for this study.

[e]Opsahl 2016a were women; Opsahl 2016b were men.

lnOR = natural log of the odds ratio; StB = standardized beta coefficient; OR = odds ratio; P = p-value; OMPQ = Orebro Musculoskeletal Pain Questionnaire; N/A = data not available or data conversions were not appropriate, but direction of association is reported; HR = hazard ratio; F = F-statistic one-way ANOVA; MD = mean difference; B = beta coefficient; SE = standard error.

Table 4. Reported associations for studies measuring important recovery outcomes

Study	Sample size	Expectation measure	Outcome measure	Outcome follow-up period	Study reported effect; variance measure	Common effect size (lnOR)[a]	Standard error	Reported direction of association[b]
Beneciuk 2017	688	General: Expectations of recovery (0 - 10, categorised in-	Non-recovery in disability (RMDQ ≥ 7)	3 months	OR; 95% CI	0.53[c]	0.24	+

Table 4. Reported associations for studies measuring important recovery outcomes *(Continued)*

		to tertiles, higher = worse)						
Besen 2015	359	General: Sum of 3 items on likely duration of symptoms (5-pt; 0 - 2 days up to > 60 days; 3 - 15)	Unresolved pain (NRS ≥ 5), function (> 50% RMDQ items endorsed), or return to work (unable to resume full duty work)	3 months	OR; 95% CI	0.19[c]	0.03	+
Downie 2016	653	General: Risk of persistence (0 - 10, higher = worse)	Belonging to a persistent pain cluster (NRS ≥ 5 at baseline and follow-up) vs rapid recovery (NRS ≤ 1 at follow-up)	3 months	RR; 95% CI	0.51	0.21	+
Enthoven 2016	422	Expectations of recovery in 3 months (1 - 5)	Belonging to a high pain trajectory cluster (consistent high pain) vs low pain trajectory cluster (0 - 1 on 10-pt NRS after 6 months)	36 months	OR; 95% CI	1.25[c]	0.26	+
Enthoven 2006	141	Treatment: Expectations of restoration (5-pt), dichotomised as quite improved/partial relief/no expectations of being restored vs completely restored	Non-recovery in disability (> 20% on ODI)	12 months	OR; 95% CI	0.17	0.54	Ø
Foster 2008	806	General: Revised Illness Perceptions Questionnaire, timeline acute/chronic item (5-pt Likert), dichotomised as least helpful perceptions (lower quartile) vs most helpful perceptions (upper quartile)	Non-recovery in disability (< 30% change in RMDQ)	6 months	RR; 95% CI	0.61	0.15	+
Harkapaa 1996	175	General: Health Optimism Index (5 - 20); reverse-coded	Positive change in disability at follow-up (≥ 3 increase in FCI disability score)	12 months	OR; 95% CI	0.20	0.07	+

Table 4. Reported associations for studies measuring important recovery outcomes *(Continued)*

		for analyses (higher = better)						
Henschke 2008	969	General: Perceived risk of persistence (0 - 10, higher = worse)	Complete recovery (pain/function/return to work)	12 months	HR; 95% CI	0.08	0.02	+
Michaelson 2004	129	Optimism Index (average of 10 items, unclear scale, higher = better)	Reduced pain (decrease in VAS ≥ 25)	12 months	P[d]	N/A	N/A	Ø
Niemisto 2004	196	General: Self-rated prognosis of work ability after 2 years (item from Workability index (1 - 7) dichotomised as poor vs good or moderate	Poor recovery (pain/function)	12 months	OR; 95% CI	0.75	0.44	Ø
Petersen 2007	158	General: Certainty of working 6 months after treatment (0 - 10), dichotomised at median as low vs high	Poor recovery (< 15% improvement from baseline disability in Low Back Pain Rating Scale)	14 months	OR; 95% CI	0.59	0.43	Ø
Rasmussen-Barr 2012	71	Treatment: Expectation of treatment (unclear scale); dichotomised as good/improved, similar vs not improved/not good for analyses	Poor recovery in disability (ODI ≥ 20)	12 months	OR; 95% CI	0.47[c]	0.56	Ø
Rundell 2017	4143	General: Expectation for recovery (0 - 10, higher = better)	Persistent disability (RMDQ ≥ 4)	6 and 12 months	OR; 95% CI	0.16	0.01	+

Table 4. Reported associations for studies measuring important recovery outcomes *(Continued)*

Shaw 2009	519	General: Likelihood of full return to work within 4 weeks (4-pt), dichotomised as definitely vs unlikely/not sure	Unresolved pain (NRS ≥ 5), function (> 50% RMDQ items endorsed), or return to work (unable to resume full duty work)	3 months	RR; 95% CI	0.99	0.37	+
Van Hooff 2014	524	Pain Self-Efficacy Questionnaire (0 - 60, higher = better)	Successful recovery (ODI ≤ 22)	12 months	Chi^2; P	1.68[c]	0.51	+
Yelland 2006	110	Treatment: Desired improvement in function to make treatment worthwhile (0 - 100%, 10% change required, higher = better)	Achieving minimum worthwhile reduction in disability determined at baseline (rated 0 - 100%)	12 months	OR; 95% CI	0.17	0.09	Ø

Table 4. Description of the reported associations between the primary expectations measure and important recovery outcomes, including presentation as common natural log odds effect size and standard error. Results presented are from the best adjusted multivariate model, when available, selecting the available study time period in study closest to 12 months (positive association in 10 studies; no association in 6 studies).

[a]All reported associations have been converted to the natural log odds (lnOR) scale and the same direction when possible; lnOR > 1 indicates a positive direction of association between expectations and outcome.

[b]Direction of association: + = positive, associated with better outcome; Ø = neutral, no association with outcome; - = negative, associated with worse outcome

[c]Study where results are from unadjusted models.

[d]Statistical significance only reported for this study.

lnOR = natural log of the odds ratio; RR = relative risk; OR = odds ratio; RMDQ = Roland Morris Disability Questionnaire; NRS = pain numeric rating scale; ODI = Oswestry Disability Index; FCI = Functional Capacity Index; VAS = pain visual analog scale; P = p-value; N/A = data not available or conversions were not appropriate, but direction of association is reported.

Table 5. Reported associations for studies measuring functional limitations

Study	Sample size	Expectation measure	Outcome measure	Outcome follow-up period	Study reported effect; variance measure	Common effect size (lnOR)[a]	Standard error	Reported direction of association[b]

Table 5. Reported associations for studies measuring functional limitations *(Continued)*

Bishop 2015	420	Self-efficacy: Chronic Pain Self-Efficacy for Pain Management subscale (0 - 100, higher = better)	RMDQ (0 - 24, higher = worse)	6 months	B; SE	0.07	0.02	+
Casey 2008	84	General: Pain Behaviour and Perception Inventory, pain permanence subscale (−2 to +2, higher = worse)	Pain-Disability Index (0 - 70, higher = worse)	3 months	B; SE	5.64	1.43	+
Demmelmaier 2010	256	General: Pain expectations (sum of 2 7-pt rating scales; 1 adapted from OMPQ; 0 - 12, higher = worse)	Disability score from Graded Chronic Pain Scale (0 - 30, higher = worse)	12 months	B; SE	1.04	0.39	+
Dionne 1997	490	General: Expectation of continued pain (4-pt)	RMDQ (16-item, higher = worse)	24 months	P[c]	N/A[d]	N/A	Ø
Goldstein 2002	650	Treatment: Treatment confidence, NRS (0 - 10, higher = better)	RMDQ (0 - 24, higher = worse)	6 months	B; 95% CI	0.65	0.43	Ø
Karjalainen 2003	161	General: Perceived risk of not recovering (0 – 10, 2-unit change required, 5-pt, higher = worse)	ODI (0 - 100, higher = worse)	12 months	B; 95% CI	2.21	0.34	+
Kongsted 2014	928	General: Likelihood of recovery (0 - 10, higher = better)	RMDQ (0 - 24, higher = worse)	12 months	R^2; P	N/A[d]	N/A	+
Macedo 2014	172	Self-efficacy: Pain Self-Efficacy Questionnaire (0 - 100), dichotomised at median as high vs low for analyses	Patient-Specific Functional Scale (0 - 10, higher = better)	12 months	B; 95% CI	1.11[d]	0.50	+
Morlock 2002	111	Treatment: Expected benefit from treatment (5 items, each 1 - 5; 0 - 100 reported, higher = better)	NASS scale (0 - 100, higher = worse)	12 months	B; P	14.20	5.51	+
Myers 2007	365	General: How much improvement do you expect in 6 weeks? (0 - 10, higher = better)	Improvement in RMDQ (0 - 23, higher = better)	3 months	B; 95% CI	0.59	0.20	+
Sherman 2009	638	Self-efficacy: Likelihood of self-managing future back pain (unclear scale, higher = better), dichotomised as top tertile vs low two tertiles	RMDQ (0 - 23, higher = worse)	12 months	B; SE	0.20	1.00	Ø

Table 5. Reported associations for studies measuring functional limitations *(Continued)*

Tran 2015a[e]	63	Treatment: How helpful do you expect yoga to be for your back problems? (0 - 10, higher = better)	Change in RMDQ (0 - 24, higher = better)	3 months	MD; P	−0.43[d]	0.55	Ø
Tran 2015b[e]	30	Treatment: How helpful do you expect yoga to be for your back problems? (0 - 10, higher = better)	Change in RMDQ (0 - 24, higher = better)	3 months	MD; P	−0.28[d]	0.5	Ø
Underwood 2007	700	Treatment: Treatment helpfulness (3-pt, not helpful, helpful, very helpful), very helpful vs not helpful compared here	RMDQ (0 - 24)	12 months	B; 95% CI	0.60	0.54	Ø

Table 5. Description of the reported associations between the primary expectations measure and function outcomes, including presentation as common natural log odds effect size and standard error. Results presented are from the best adjusted multivariate model, when available, selecting the available study time period in study closest to 12 months (positive association in 9 studies; no association in 5 studies (6 groups)).

[a]All reported associations have been converted to the natural log odds (lnOR) scale and the same direction when possible; lnOR > 1 indicates a positive direction of association between expectations and outcome.

[b]Direction of association: + = positive, associated with better outcome; Ø = neutral, no association with outcome; - = negative, associated with worse outcome

[c]Statistical significance only reported for this study.

[d]Study where results are from unadjusted models.

[e]Tran 2015a received twice-weekly yoga; Tran 2015b received once-weekly yoga.

lnOR = natural log of the odds ratio; RMDQ = Roland Morris Disability Questionnaire; B = beta coefficient; SE = standard error; OMPQ = Orebro Musculoskeletal Pain Questionnaire; P = p-value; N/A = data not available or data conversions were not appropriate, but direction of association is reported; NRS = pain numeric rating scale; ODI = Oswestry Disability Index; NASS = North American Spine Society scale; MD = mean difference.

Table 6. Reported associations for studies measuring pain intensity outcomes

Study	Sample size	Expectation measure	Outcome measure	Outcome follow-up period	Study reported effect; variance measure	Common effect size (lnOR)[a]	Standard error	Reported direction of association[b]
Casey 2008	84	General: Pain Behaviour and Perception Inventory, pain permanence subscale (−2 to +2, higher = worse)	Descriptor Differential Scale (0 - 20, higher = worse)	3 months	B; SE	0.86	0.56	Ø
Demmelmaier 2010	256	General: Pain expectations (sum of 2 7-pt rating scales; 1	Pain scale from Graded Chronic Pain Scale (0 - 30, higher = worse)	12 months	B; SE	0.95	0.35	+

Table 6. Reported associations for studies measuring pain intensity outcomes *(Continued)*

		adapted from OMPQ; 0 - 12, higher = worse)						
Glattacker 2013	81	General: Revised Illness Perceptions Questionnaire, timeline acute/chronic item (5-pt Likert, higher = worse)	LBP Intensity VAS (0 - 100, higher = worse)	6 months	St B; P	0.24	0.10	+
Goldstein 2002	650	Treatment: Treatment confidence NRS (0 - 10, higher = better)	LBP Intensity Change in NRS (0 - 10, higher = worse)	6 months	B; 95% CI	0.28	0.19	Ø
Haas 2014	391	Treatment: Confidence in treatment, average of 2 6-pt Likert scales (1 - 6, higher = better)	Von Korff (0 - 100, higher = worse)	3 months	B; 95% CI	0.05	0.04	Ø
Jensen 2000	107	Self-efficacy: Belief in ability for learning to cope with the pain (unclear scale, higher = better)	SF-36 Bodily Pain Scale (0 - 100, higher = better)	6 months	B; 95% CI	9.44	3.94	+
Karjalainen 2003	161	General: Perceived risk of not recovering (0 – 10, 2-unit change required, 5-pt, higher = worse)	LBP Intensity NRS (0 - 10, higher = worse)	6 months	B; 95% CI	0.32	0.12	+
Kongsted 2014**a**[c]	200	General: Likelihood of recovery (0 - 10, higher = better)	LBP Intensity NRS (0 - 10, higher = worse)	3 months	B; 95% CI	0.19	0.07	+
Kongsted 2014**b**[c]	705	General: Likelihood of recovery (0 - 10, higher = better)	LBP Intensity NRS (0 - 10, higher = worse)	3 months	B; 95% CI	0.12	0.04	+
Macedo 2014	172	Self-efficacy: Pain Self-Efficacy Questionnaire (0 - 100), dichotomised at median as high vs low for analyses	LBP Intensity NRS (0 - 10, higher = worse)	12 months	B; 95% CI	0.92[d]	0.58	Ø
Tran 2015**a**[e]	63	Treatment: How helpful do you expect yoga to be for your back problems? (0 - 10, higher = better)	Change in NRS (0 - 10, higher = better)	3 months	MD; P	−0.20[d]	0.45	Ø
Tran 2015**b**[e]	30	Treatment: How helpful do you expect yoga to be for your back	Change in NRS (0 - 10, higher = better)	3 months	MD; P	−0.43[d]	0.44	Ø

Table 6. Reported associations for studies measuring pain intensity outcomes *(Continued)*

	problems? (0 - 10, higher = better)

Table 6. Description of the reported associations between the primary expectations measure and pain intensity outcomes, including presentation as common natural log odds effect size and standard error. Results presented are from the best adjusted multivariate model, when available, selecting the available study time period in study closest to 12 months (positive association in 5 studies (6 groups); no association in 5 studies (6 groups)).

[a]All reported associations have been converted to the natural log odds (lnOR) scale and the same direction when possible; lnOR > 1 indicates a positive direction of association between expectations and outcome.

[b]Direction of association: + = positive, associated with better outcome; Ø = neutral, no association with outcome; - = negative, associated with worse outcome

[c]Kongsted 2014a was a general practice cohort; Kongsted 2014b was a chiropractic practice cohort.

[d]Study where results are from unadjusted models.

[e]Tran 2015a received twice-weekly yoga; Tran 2015b received once-weekly yoga.

lnOR = natural log of the odds ratio; B = beta coefficient; SE = standard error; LBP = low back pain; VAS = pain visual analogue scale; StB = standardized beta coefficient; P = p-value; NRS = pain numeric rating scale; SF-36 = 36-item Short Form survey; MD = mean difference.

Table 7. QUIPS Risk of bias domain summary by study

Study	Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting
Beneciuk 2017	2017	Moderate	Moderate	Moderate	Low	High	Low
Besen 2015	2015	Moderate	Moderate	Low	Low	Low	Low
Bishop 2015	2015	High	Moderate	Low	Low	Low	Low
Butler 2007	2007	Moderate	Moderate	Moderate	Low	Low	Low
Carriere 2015	2015	Low	Moderate	Moderate	Moderate	Low	Low
Casey 2008	2008	Moderate	Moderate	Low	Low	Moderate	Low
Demmelmaier 2010	2010	High	High	High	Low	Moderate	Moderate
Dionne 1997	1997	Moderate	Low	Moderate	Low	High	Low
Downie 2016	2016	Moderate	Low	Low	Moderate	Low	Low
Du Bois 2008	2008	Moderate	Low	Moderate	Low	Moderate	Moderate

Table 7. QUIPS Risk of bias domain summary by study *(Continued)*

Enthoven 2006	2006	Low	Low	Moderate	Low	Low	Low
Enthoven 2016	2016	Moderate	Low	Low	Low	High	Moderate
Foster 2008	2008	Moderate	High	Low	Low	Low	Low
George 2010	2010	Low	Low	Moderate	Low	High	Low
Gervais 1991	1991	Low	Low	Low	Low	Moderate	Low
Glattacker 2013	2013	Moderate	Moderate	Low	Low	Low	Low
Goldstein 2002	2002	Moderate	Low	Low	Low	Moderate	Low
Gross 2010	2010	Moderate	Low	Low	Low	Low	Low
Grotle 2006	2006	High	Moderate	Moderate	Low	Moderate	Low
Haas 2014	2014	High	Low	Low	Low	Moderate	Low
Hagen 2005	2005	Moderate	Low	Moderate	Low	Low	Low
Haldorsen 1998	1998	Moderate	Low	High	Low	High	Moderate
Harkapaa 1996	1996	High	Moderate	Low	Low	Low	Low
Hazard 1996	1996	Moderate	Low	Low	Low	Moderate	Moderate
Henschke 2008	2008	Low	Low	Low	Low	Low	Low
Heymans 2006	2009	Moderate	Low	Low	Low	Low	Low
Hildebrandt 1997	1997	High	Moderate	Moderate	Low	High	Moderate
Jellema 2002	2002	High	Moderate	Low	Low	High	High
Jensen 2000	2000	Moderate	Moderate	Moderate	Low	Moderate	Low
Jensen 2013	2013	Moderate	Low	Moderate	Low	Moderate	Low
Karjalainen 2003	2003	Moderate	Low	Low	Low	Low	Moderate

Table 7. QUIPS Risk of bias domain summary by study *(Continued)*

Kongsted 2014	2014	Low	Moderate	Low	Low	Low	Low
Leboeuf-Yde 2004	2004	Moderate	Moderate	Moderate	Low	Moderate	Moderate
Lindell 2010	2010	Low	Low	Moderate	Low	Low	Low
Macedo 2014	2014	Moderate	Low	Moderate	Low	High	Low
Magnussen 2007	2007	Moderate	Moderate	Moderate	Moderate	High	Low
Michaelson 2004	2004	Low	Low	Low	Low	Moderate	Moderate
Morlock 2002	2002	Moderate	High	Low	Low	Low	Moderate
Myers 2007	2007	Low	Low	Low	Low	Low	Low
Niemisto 2004	2004	Low	Low	Moderate	Low	Low	Moderate
Opsahl 2016	2016	Low	Low	Moderate	Moderate	Low	Low
Opsommer 2017	2017	Low	Moderate	Low	Moderate	High	Low
Petersen 2007	2007	Low	Low	Moderate	Low	Moderate	Low
Rasmussen-Barr 2012	2012	Low	Moderate	Moderate	Low	High	Low
Reeser 2001	2001	Moderate	High	Low	Low	High	Moderate
Reiso 2003	2003	Moderate	Low	Moderate	Low	Low	Moderate
Reme 2009	2009	Low	Low	Moderate	Low	Low	Moderate
Rundell 2017	2017	Moderate	Moderate	Low	Low	Moderate	Low
Sandstrom 1986	1986	Moderate	Low	Moderate	Low	High	Moderate
Schultz 2004	2004	High	Low	Low	Low	Moderate	Moderate
Shaw 2009	2009	High	Moderate	Moderate	Moderate	Low	Low
Sherman 2009	2009	Moderate	Low	Moderate	Low	Low	Low

Table 7. QUIPS Risk of bias domain summary by study *(Continued)*

Steenstra 2005	2005	Low	Low	Moderate	Low	Moderate	Low
Tran 2015	2015	Moderate	Low	Moderate	Low	Moderate	Moderate
Truchon 2012	2012	Moderate	Low	High	Low	Low	Low
Turner 2008	2008	Moderate	Low	Moderate	Low	Low	Low
Underwood 2007	2007	Low	Moderate	Moderate	Low	Low	Low
Van Hooff 2014	2014	Low	Low	Low	Low	Moderate	Moderate
Van Wijk 2008	2008	Moderate	Low	High	Low	Low	Low
Yelland 2006	2006	Moderate	Low	Low	Low	Moderate	Low

Table 8. Detailed QUIPS risk of bias assessments by study

Study ID: Beneciuk 2017		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate 63%; non-participants not adequately described
Study Attrition	Moderate	80% follow-up at 4 months, 76% at 1 year; dropouts were younger
Prognostic Factor Measurement	Moderate	PF with good face validity (expectations of recovery; 0 - 10); data driven cut-points used to categorise continuous measure of PF
Outcome Measurement	Low	Clinical rationale provided for categorisation of continuous outcome (RMDQ)
Study Confounding	High	Univariate only for PF association with outcome (TEM analysis)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Besen 2015		
Domain	**Risk of bias level**	**Support for Judgement**
Study Participation	Moderate	Participation rate unclear (volunteers); selection criteria and baseline sample adequately described
Study Attrition	Moderate	72% follow-up at 3 months; no reasons for loss, respondents had more organisational support, which may bias results
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (RTW confidence, and RTW self-efficacy scale)
Outcome Measurement	Low	Valid and reliable measurement of outcome (VAS, Quebec Back Pain Disability scale); work status, work modifications, duration of absences self-reported with unclear measurement properties
Study Confounding	Low	Adequate adjustment (education, fear avoidance, catastrophising, race, ethnicity, income, organisational support, co-worker support, pain)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results; conceptual framework for inclusion of covariates
Study ID: Bishop 2015		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	High	Participation rate 38%; non-participants not adequately described
Study Attrition	Moderate	87% follow-up; no information on attempts to collect outcome information from dropouts; dropouts were younger
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (subscale of the Credibility Expectancy Questionnaire)

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Outcome Measurement	Low	Valid and reliable measure of outcome (RMDQ)
Study Confounding	Low	Adequate adjustment (age, work status, LBP-related benefits status, LBP-related compensation status, reporting at least 1 comorbidity, reporting at least 1 co-treatment, duration of LBP, clinic type, and healthcare sector)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Butler 2007		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate 51%; non-participants not adequately described
Study Attrition	Moderate	87% follow-up at 1 month, 62% at 6 months, 42% at 1 year; no information on attempts to collect or possible reasons
Prognostic Factor Measurement	Moderate	Continuous measure of PF with good face validity (5-point expectations of recovery) dichotomised for analyses without rationale provided
Outcome Measurement	Low	Unclear validity and reliability measurement of RTW outcome (unstable employment pattern)
Study Confounding	Low	Adequate adjustment (age, work status, LBP-related benefits status, LBP-related compensation) status, reporting at least 1 comorbidity, reporting at least 1 co-treatment, duration of LBP, clinic type, and healthcare sector)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Carriere 2015		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Low	Participants were consecutive referrals to a clinic
Study Attrition	Moderate	78% follow-up; no description of reasons
Prognostic Factor Measurement	Moderate	Data driven cut-point used to categorise a continuous measure of PF with good face validity (likelihood of RTW; 0 - 100)
Outcome Measurement	Moderate	Unclear measure of RTW status
Study Confounding	Low	Adeqaute adjustment (age, sex, work disability, pain severity, number of pain sites)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Casey 2008		
Domain	**Risk of bias level**	**Support for judgement**

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Study Participation	Moderate	Participation rate unclear; selection criteria and baseline characteristics adequately described
Study Attrition	Moderate	87% follow-up; significant difference in those lost to follow-up on pain constancy
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (Pain Behaviour and Perception Inventory pain permanence subscale)
Outcome Measurement	Low	Valid and reliable measurement of outcome (mean VAS, PDI)
Study Confounding	Moderate	Minimal adjustment (previous pain, baseline pain, baseline disability, cumulative trauma, depressive symptoms, pain permanence beliefs, pain constancy beliefs, chronic pain intensity (3 months))
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Demmelmaier 2010		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	High	Participation rate 39%; non-participants not adequately described
Study Attrition	High	37% follow-up; no information provided on reasons or differences in characteristics
Prognostic Factor Measurement	High	Unclear measurement properties of PF; group median used to substitute missing data
Outcome Measurement	Low	Valid and reliable measurement of outcome (GCPS, sick leave yes/no)
Study Confounding	Moderate	Minimal adjustment (pain intensity baseline, disability, pain catastrophising, fear of movement, functional self-efficacy)
Statistical Analysis and Reporting	Moderate	Appropriate analysis for research question and study design; possible selective reporting of results
Study ID: Dionne 1997		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate 72%; non-participants not adequately described
Study Attrition	Low	92% follow-up
Prognostic Factor Measurement	Moderate	Unclear validity and measurement properties (5-point expectation of continued pain with no time period included)
Outcome Measurement	Low	Valid and reliable measurement of outcome (RMDQ)
Study Confounding	High	Univariate only available for our review question

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations

Study ID: Downie 2016

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate 67%; non-participants not adequately described
Study Attrition	Low	96% follow-up at 3 months
Prognostic Factor Measurement	Low	Continuous measure of PF with good face validity (risk of persistence; 0 - 10)
Outcome Measurement	Moderate	Persistent high pain (≥ 5 on NRS) cluster and rapid recovery (< 1 on NRS) data available (clusters represent extremes)
Study Confounding	Low	Adequate adjustment (age, sex, taking paracetamol, compensable, pain intensity, duration, pain beyond knee, previous episodes, days of reduced activity, poor sleep quality, quality of life physical, quality of life mental, disability)
Statistical Analysis and Reporting	Low	While we were only able to use a subset of the data, authors used an appropriate analysis for research question and study design; no apparent selective reporting of results

Study ID: Du Bois 2008

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate unclear; selection criteria and baseline sample adequately described
Study Attrition	Low	100% follow-up; work status recorded by the sickness fund
Prognostic Factor Measurement	Moderate	Continuous measure of PF with good face validity (10-point Likert scale for probability of RTW within 6 months); dichotomised for analyses without rationale
Outcome Measurement	Low	Valid and reliable measurement of outcome (RTW at 3 months)
Study Confounding	Moderate	Minimal adjustment (pain below knee, pain interference)
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations

Study ID: Enthoven 2006

Domain	Risk of bias level	Support for judgement
Study Participation	Low	Participation rate 58%; inclusion criteria and non-participants adequately described
Study Attrition	Low	93% follow-up at 1 year, 83% at 5 years
Prognostic Factor Measurement	Moderate	Continuous measure of PF with good face validity (5-point expectations of restoration) dichotomised for analyses without rationale

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Outcome Measurement	Low	Valid and reliable measurement of outcome (ODI, sick leave duration)
Study Confounding	Low	Adequate adjustment (age, sex, duration of current episode, similar problems during previous 5 years, exercise frequency before, exercise level before, dissatisfied with work, dissatisfied with workplace, more than one localisation, pain frequency, ODI score, well-being, current sick leave)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Enthoven 2016		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate 53%; non-participants not adequately described
Study Attrition	Low	93% follow-up at 3 months; reasons for loss provided
Prognostic Factor Measurement	Low	Continuous measure of PF with good face validity (5-point expectations of recovery)
Outcome Measurement	Low	Valid and reliable measurement of outcome (NRS, RMDQ)
Study Confounding	High	Univariate only available for meta-analyses
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations
Study ID: Foster 2008		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate 51.4%; non-participants not adequately described
Study Attrition	High	65% follow-up at 6 months; no information provided on reasons for loss
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (IPQ-R items) (Moss-Morris 2002)
Outcome Measurement	Low	Valid and reliable measurement of outcome (pain duration); clinical rationale used to dichotomise RMDQ
Study Confounding	Low	Adequate adjustment (sex, education, catastrophising, fear avoidance, social class, pain intensity, RMDQ, pain duration, leg pain, distal pain, anxiety, depression, 4 more domains of CSQ, 6 domains of IPQR, passive coping)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: George 2010		
Domain	**Risk of bias level**	**Support for judgement**

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Study Participation	Low	Participation rate 81%
Study Attrition	Low	67% follow-up at 6 months; no differences found between dropouts and those with follow-up data
Prognostic Factor Measurement	Moderate	Valid and reliable measure of PF (items from NASS lumbar spine outcome assessment instrument); dichotomised for analyses without rationale
Outcome Measurement	Low	Valid and reliable measure of outcome (von Korff)
Study Confounding	High	Univariate only
Statistical Analysis and Reporting	Low	Analysis not sufficient for our review question; combines 4-week and 6-month outcome data
Study ID: Gervais 1991		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Low	Participation rate 96%
Study Attrition	Low	98% follow-up
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (self-efficacy and results expectancies inventory)
Outcome Measurement	Low	Valid and reliable measurement of outcome (RTW time, recurrence, and at 6 months)
Study Confounding	Moderate	Minimal adjustment (diagnosis, lowest pain intensity, length of inactivity before treatment, negative life changes)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Glattacker 2013		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate 59%; non-participants not adequately described
Study Attrition	Moderate	74% follow-up; no description of reasons; dropouts had higher baseline disability may bias results
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (IPQ-R items) (Moss-Morris 2002)
Outcome Measurement	Low	Valid and reliable measurement of outcome (VAS, ODI, SF-36 scales)
Study Confounding	Low	Adequate adjustment (baseline health, baseline mental health, age, sex, education, illness duration, psychological outcome expectation, process expectation, rehabilita-

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

		tion-specific concerns, identify, consequences, personal control, treatment control, coherence, emotional representation, cause (overwork))
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Goldstein 2002		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate 46.4%; reasons for not participating adequately described
Study Attrition	Low	96% follow-up
Prognostic Factor Measurement	Low	Continuous measure of PF with good face validity (NRS treatment confidence, 0 - 10)
Outcome Measurement	Low	Valid and reliable measurement of outcome (NRS, RMDQ)
Study Confounding	Moderate	Minimal adjustment (baseline disability, age, gender, race, treatment group, duration of LBP episode)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Gross 2010		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participants drawn from workers' compensation database; non-participants not described
Study Attrition	Low	Claim-based outcome measure; available for all participants included
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (Work-related Recovery Expectations Questionnaire)
Outcome Measurement	Low	Valid and reliable measurement of outcome (time to claim closure or suspension of time-loss benefits, recurrence)
Study Confounding	Low	Adequate adjustment (age, sex, job attachment status, duration of injury, number of previous claims, urban/rural)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Grotle 2006		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	High	Participation rate unclear; supplemental recruiting from general population
Study Attrition	Moderate	91% and 94% follow-up at 1 year for each group; no reasons provided for loss to follow-up

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Prognostic Factor Measurement	Moderate	Valid and reliable measure of PF (Acute Low Back Pain Screening Questionnaire item), data driven cut-point used for analyses
Outcome Measurement	Low	Valid and reliable measurement of outcome (NRS, days of participation restriction)
Study Confounding	Moderate	Minimal adjustment (age, sex)
Statistical Analysis and Reporting	Low	Analysis not sufficient for our study purposes only, expectations included as an item of a larger measure
Study ID: Haas 2014		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	High	Participation rate 42%; recruitment from general population
Study Attrition	Low	98% follow-up
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (Interstudy's Low Back Pain TyPE Specification)
Outcome Measurement	Low	Valid and reliable measure of outcome (von Korff)
Study Confounding	Moderate	Minimal adjustment for relationship between baseline expectations and outcome (expectations at 6 and 12 weeks, doctor-participant encounter at 6 and 12 weeks, LBP at 6 and 12 weeks)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design with conceptual model
Study ID: Hagen 2005		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate unclear; selection criteria and baseline characteristics adequately described
Study Attrition	Low	99% follow-up at 1 year
Prognostic Factor Measurement	Moderate	Unclear validity, reliability of PF measurement and unclear measurement properties (Do you believe back pain will disappear?)
Outcome Measurement	Low	RTW determined from administrative claims database
Study Confounding	Low	Adequate adjustment (gender, age, education, group (intervention vs control), large reduced ability to regularly work, constant back strain > 50% of the working time, gastrointestinal problems, high chance externality (health locus of control), believe work will aggravate condition, pain when performing daily activities, state anxiety, other illnesses + 4 interaction terms)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Study ID: Haldorsen 1998

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate 54%; non-participants not adequately described; baseline sample adequately described
Study Attrition	Low	100% follow-up at 1 year
Prognostic Factor Measurement	High	Unclear validity, reliability of PF and unclear measurement properties of PF (If you continue working, what effect will that have on your complaints?)
Outcome Measurement	Low	Valid and reliable measurement of outcome (sick-listed)
Study Confounding	High	Univariate only
Statistical Analysis and Reporting	Moderate	Appropriate analysis for research question and study design; possible selective reporting of results

Study ID: Harkapaa 1996

Domain	Risk of bias level	Support for judgement
Study Participation	High	Participation rate unclear; no description of non-participants, limited description of participants
Study Attrition	Moderate	74% follow-up at 1 year; no information provided on reasons or differences in characteristics
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (Optimism Index)
Outcome Measurement	Low	Valid and reliable measurement of outcome (work status), valid rationale used to categorise Functional Capacity Index
Study Confounding	Low	Adequate adjustment (age, sex, work status at baseline, baseline disability, others' locus of control)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results

Study ID: Hazard 1996

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate 24%; non-participants were described and compared with small mean differences
Study Attrition	Low	98% follow-up at 3 months
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (Vermont Disability Prediction Questionnaire item)
Outcome Measurement	Low	Self-report measure of RTW with unclear measurement properties

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Study Confounding	Moderate	Insufficient data on which Vermont Disability Prediction Questionnaire domains included in final multivariate model; age and sex not included
Statistical Analysis and Reporting	Moderate	Analysis not sufficient for our study purposes only, expectations included as an item of a larger measure; no conceptual framework; data driven based on P values of univariate associations
Study ID: Henschke 2008		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Low	Participation rate 92.1%
Study Attrition	Low	99% follow-up
Prognostic Factor Measurement	Low	Continuous measure of PF with good face validity (NRS perceived risk of persistence, 0 - 10)
Outcome Measurement	Low	Valid and reliable measurement of outcome (SF-36, return to previous work status)
Study Confounding	Low	Adequate adjustment (age, sex, pain intensity, interference with function, pain control, tension/anxiety, feelings of depression, compensable LBP, currently taking medications for LBP, days of reduced activity due to LBP, leg pain, no of pain sites, duration of episodes)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Heymans 2006		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate unclear; selection criteria and baseline sample adequately described
Study Attrition	Low	100% follow-up (90% with complete baseline data)
Prognostic Factor Measurement	Low	Good face validity, unclear reliability of PF (time to RTW in categories and 5-point scale for certainty of RTW)
Outcome Measurement	Low	Valid and reliable measurement of outcome (time to full RTW)
Study Confounding	Low	Adequate adjustment (job satisfaction, social support, pain radiation in 1 or both legs, pain intensity)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; comprehensive method for backwards selection of variables
Study ID: Hildebrandt 1997		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	High	Participation rate and recruitment approach unclear

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Study Attrition	Moderate	91% follow-up; some non-participants lost due to refusal to participate, and not adequately described
Prognostic Factor Measurement	Moderate	Mix of treatment and general expectations; unclear measurement properties of PF (RTW after treatment)
Outcome Measurement	Low	Valid and reliable measurement of outcome (VAS, sick-listed Y/N)
Study Confounding	High	Univariate only
Statistical Analysis and Reporting	Moderate	Analysis not sufficient for our study purposes only, outcome measured at 6 and 12 months but not analysed; Only mean and standard deviation of expectations measure presented for success and failure in outcome

Study ID: Jellema 2002

Domain	Risk of bias level	Support for judgement
Study Participation	High	Participation rate unclear (workers volunteered)
Study Attrition	Moderate	83% follow-up; reasons for loss to follow-up provided, but no information provided on differences in characteristics
Prognostic Factor Measurement	Low	Continuous measure of PF with good face validity (NRS treatment confidence, 0 - 10)
Outcome Measurement	Low	Valid and reliable measurement of outcome (NRS, Quebec Back Pain Disability Scale, time lost from work)
Study Confounding	High	Univariate data only measures change in pain intensity with no other baseline variables taken into account; no data available for meta-analyses
Statistical Analysis and Reporting	High	Analysis not sufficient for our study purposes only, outcome measured weekly and mean benefit over 6 months used for analysis

Study ID: Jensen 2000

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate unclear; selection criteria and baseline characteristics adequately described
Study Attrition	Moderate	69% follow-up in sample 1, 100% in sample 2; no information provided on reasons or differences in characteristics
Prognostic Factor Measurement	Moderate	Unclear measurement properties of PF; unclear validity, reliability of PF (Belief that there is a treatment that could relieve condition; Belief in ability for learning to cope with the pain)
Outcome Measurement	Low	Valid and reliable measurement of outcome (SF-36, days sick leave, disability pension Y/N)

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Study Confounding	Moderate	Minimal adjustment (attending physician's judgement, attending physiotherapist's judgement of need and potential; insurance officer judgement of need and potential; screening physician judgement of need and of potential)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results

Study ID: Jensen 2013

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate unclear; selection criteria and baseline characteristics adequately described
Study Attrition	Low	100% follow-up
Prognostic Factor Measurement	Moderate	Good face validity, unclear reliability of PF (time to RTW categories)
Outcome Measurement	Low	Valid and reliable measurement of outcome (successful RTW for at least 4 weeks)
Study Confounding	Moderate	Minimal adjustment (pain side flexion, blaming work for pain, drinking alcohol, radiculopathy, BMI, age, sex)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; method for backwards selection of variables

Study ID: Karjalainen 2003

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate unclear; selection criteria and baseline characteristics adequately described
Study Attrition	Low	95% follow-up
Prognostic Factor Measurement	Low	Continuous measure of PF with good face validity (NRS expectations of treatment effectiveness, 0 - 10)
Outcome Measurement	Low	Valid and reliable measurement of outcome (NRS, ODI, sick leave because of back pain)
Study Confounding	Low	Adequate adjustment (gender, age, BMI, blue-collar worker, duration of sick leave at baseline, radicular symptoms below the knee, intensity of pain at baseline, ODI, satisfaction with work, self-rated health status for age)
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations

Study ID: Kongsted 2014

Domain	Risk of bias level	Support for judgement
Study Participation	Low	Participation rate 98%

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Study Attrition	Moderate	Unclear attrition at 6 months and 1 year, 76% - 83% follow-up at 3 months; no information provided on reasons or differences in characteristics
Prognostic Factor Measurement	Low	Continuous measure of PF with good face validity (NRS)
Outcome Measurement	Low	Valid and reliable measure of outcomes (NRS and RMDQ)
Study Confounding	Low	Adequate adjustment (Model 1: age, gender, education, any sick leave previous month, LBP at baseline, leg pain at baseline, activity limitation at baseline, duration of LBP, number previous episodes of LBP, depression)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Leboeuf-Yde 2004		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate unclear; selection criteria and baseline characteristics adequately described
Study Attrition	Moderate	68% follow-up at 3 months, 59% at 1 year; gender differences at each follow-up from baseline, no reasons for loss reported
Prognostic Factor Measurement	Moderate	Validity, reliability and measurement properties of PF unclear (helpfulness of treatment, sick leave in 6 weeks, yes/no; 4-point improvement of pain in 6 weeks)
Outcome Measurement	Low	Valid and reliable measurement of outcome (pain 0 - 10, ODI), clinical rationale used to dichotomise for analyses
Study Confounding	Moderate	Minimal adjustment at 3 months (sex, pain severity, episode duration, neck pain) and 12 months (disability, activity limitation)
Statistical Analysis and Reporting	Moderate	Analysis not sufficient for our study purposes only, combines 4-week, 3-month and 12-month outcome data; no conceptual framework; data driven based on P values of univariate associations
Study ID: Lindell 2010		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Low	Participation rate 85%
Study Attrition	Low	99% follow-up at 6 months, 98% at 1 year
Prognostic Factor Measurement	Moderate	Continuous measure of PF with good face validity (5-point expectations of RTW) dichotomised for analyses without rationale provided
Outcome Measurement	Low	Valid and reliable measurement of outcome (successful RTW for at least 1 month)
Study Confounding	Low	Adequate adjustment (age, education, type of back pain, back pain domination, catastrophising, total prior sick-listing)

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; method for backwards selection of variables

Study ID: Macedo 2014

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate 68.5%; non-participants not adequately described
Study Attrition	Low	87% - 93% follow-up; no differences found between dropouts and those with follow-up data
Prognostic Factor Measurement	Moderate	Valid and reliable measure of expectations (PSEQ) dichotomised for analyses without rationale provided
Outcome Measurement	Low	Valid and reliable measure of outcomes (NRS and Patient-Specific Function Scale)
Study Confounding	High	Univariate only for PF association data (TEM analysis)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results

Study ID: Magnussen 2007

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate unclear; adequate description of non-participants but no reasons for not participating/exclusion
Study Attrition	Moderate	91% follow-up; reasons provided for not completing intervention but not questionnaire, no information on differences in characteristics
Prognostic Factor Measurement	Moderate	Unclear validity, reliability, and measurement of PF (ever RTW, "don't know" grouped with "no" for analyses)
Outcome Measurement	Moderate	Unclear validity and reliability measurement of RTW outcome (self-report of being in a RTW process with unclear measurement properties; additional outcome - RMDQ)
Study Confounding	High	Univariate only
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results

Study ID: Michaelson 2004

Domain	Risk of bias level	Support for judgement
Study Participation	Low	Participation rate 98%
Study Attrition	Low	75% follow-up at 1 year; authors indicate no differences between dropouts and those with follow-up data

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Prognostic Factor Measurement	Low	Valid and reliable measure of expectations (Optimism Index)
Outcome Measurement	Low	Clinical rationale provided for categorisation of continuous outcome (VAS)
Study Confounding	Moderate	Minimal adjustment (age, sex, optimism index, MPI pain severity, pain intensity on average); no data available for meta-analyses
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations
Study ID: Morlock 2002		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate unclear; inclusion criteria adequately described, baseline sample not adequately described
Study Attrition	High	Unclear attrition; no information provided on reasons or differences in characteristics
Prognostic Factor Measurement	Low	Continuous PF measure with good face validity (combining 5 5-point scales of treatment expectation)
Outcome Measurement	Low	Valid and reliable measurement of outcome (NASS)
Study Confounding	Low	Adequate adjustment (age, baseline NASS score, workers compensation status, gender, race, acuity, symptom location, Charlson Comorbidity Index)
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations
Study ID: Myers 2007		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Low	Participation rate 66%; selection criteria, reasons for not participating, and baseline characteristics adequately described
Study Attrition	Low	99% at follow-up
Prognostic Factor Measurement	Low	Continuous measure of PF with good face validity (NRS likelihood of recovery, 0 - 10)
Outcome Measurement	Low	Valid and reliable measurement of outcome (RMDQ)
Study Confounding	Low	Adequate adjustment (age, race, income, baseline disability, depression, history of sciatica, 2nd time seeing doctor for LBP, baseline pain)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; comprehensive method for backwards selection of variables
Study ID: Niemisto 2004		
Domain	**Risk of bias level**	**Support for judgement**

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Study Participation	Low	Participation rate 97%
Study Attrition	Low	96% follow-up
Prognostic Factor Measurement	Moderate	Independent measurement properties of PF unknown (item of Workability Index) dichotomised for analyses without rationale provided
Outcome Measurement	Low	Valid and reliable measurement of outcome (VAS, ODI, Work-ability Index, number of days on sick leave)
Study Confounding	Low	Adequate adjustment (university education; VAS score; sick leave days during previous year; life control; SLUMP test)
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations
Study ID: Opsahl 2016		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Low	Participation rate 100%
Study Attrition	Low	99% follow-up
Prognostic Factor Measurement	Moderate	Low and moderate expectancies grouped for analyses due to data constraints (4-point rating of extent of RTW)
Outcome Measurement	Moderate	Unclear measure of RTW status
Study Confounding	Low	Adequate adjustment (Model 1: age, education, fear avoidance, smoking status, intervention groups, subjective health complaint inventory total, ODI, Hopkins Symptom Checklist (HSCL-25) (emotional distress), co-worker social support)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Opsommer 2017		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Low	Participation rate 89.5%
Study Attrition	Moderate	79.5% follow-up; coping strategies and fear avoidance higher (worse) in dropouts may bias results
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (Orebro Musculoskeletal Pain Questionnaire items)
Outcome Measurement	Moderate	Unclear measure of RTW status
Study Confounding	High	Univariate only

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results

Study ID: Petersen 2007

Domain	Risk of bias level	Support for judgement
Study Participation	Low	Participants were consecutive referrals to a clinic; sample adequately described
Study Attrition	Low	93% follow-up at 1y, 83% at 5 years
Prognostic Factor Measurement	Moderate	Continuous measure of PF with good face validity (11-point box scale, certainty of RTW) dichotomised for analyses without rationale provided
Outcome Measurement	Low	Valid and reliable measurement of outcome (sick-listed), valid rationale used to categorise Low Back Rating Scale for analyses
Study Confounding	Moderate	Minimal adjustment (baseline disability, on sick leave, low job satisfaction, pain below the knee)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; comprehensive method for backwards selection of variables

Study ID: Rasmussen-Barr 2012

Domain	Risk of bias level	Support for judgement
Study Participation	Low	Participation rate 93%
Study Attrition	Moderate	86% follow-up at 1 year; reasons provided, lost to follow-up had lower physical health, may bias results
Prognostic Factor Measurement	Moderate	Unclear validity, reliability of PF and unclear measurement properties of PF
Outcome Measurement	Low	Valid and reliable measurement of outcome (VAS, ODI), valid rationale used to dichotomise scores for analyses
Study Confounding	High	Univariate only
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations

Study ID: Reeser 2001

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate unclear (recruitment at appointment booking); selection criteria and baseline characteristics adequately described
Study Attrition	High	35% completed questionnaires at all follow-up times at were included; respondents were older, may bias results; no reasons for loss provided

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Prognostic Factor Measurement	Low	Continuous PF measure with good face validity (combining 5 5-point scales of treatment expectation)
Outcome Measurement	Low	Valid and reliable measurement of outcome (MODEMS Physical Health and Pain scale, MODEMS Physical Health and Disability scale)
Study Confounding	High	No statistical models were used to assess the association between the expectations and the outcomes of interest
Statistical Analysis and Reporting	Moderate	No statistical models were used to assess the association between the expectations and the outcomes of interest; possible selective and unclear reporting of results
Study ID: Reiso 2003		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate unclear (referred from primary care); selection criteria and baseline sample adequately described
Study Attrition	Low	Primary outcome return to work for 60+ days supplied by national register; no loss to follow-up
Prognostic Factor Measurement	Moderate	Measure of PF with good face validity (RTW, part-time RTW, still on sick leave) dichotomised for analyses without rationale provided
Outcome Measurement	Low	Valid and reliable measure of RTW (working for at least 60 consecutive days) from national register
Study Confounding	Low	Adequate adjustment (age, gender, diagnosis, pain intensity, workability)
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations
Study ID: Reme 2009		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Low	Participation rate 65%; inclusion criteria, reasons for not participating, and baseline sample adequately described
Study Attrition	Low	99% follow-up at 3 months, 97% at 1 year
Prognostic Factor Measurement	Moderate	Unclear validity, reliability of PF measurement and unclear measurement properties of PF (Expectation to return to work within the next few weeks (yes/no))
Outcome Measurement	Low	Valid and reliable measurement of outcome (sick leave and non-RTW from insurance claims, self-reported sick-listing)
Study Confounding	Low	Adequate adjustment (group, gender, age, education, workload, sleep problems, reduced ability to walk, physiotherapy, back pain intensity, pain during activity, pain during rest)
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations
Study ID: Rundell 2017		

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Unclear participation rate; baseline sample and selection criteria adequately described; unclear reasons for not participating
Study Attrition	Moderate	79% follow-up at both 6 months and 1 year; no information provided on reasons for differences in characteristics
Prognostic Factor Measurement	Low	Continuous measure of PF at 6- and 12-month follow-up with good face validity (confidence in recovery; 0 - 10); categorised for 3-month follow-up
Outcome Measurement	Low	Valid and reliable measure of outcomes (NRS and RMDQ); clinical rationale provided for categorisation of continuous
Study Confounding	Moderate	Minimal adjustment (age, sex, education, race, employment status, marital status, study site)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Sandstrom 1986		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Reported as consecutive referrals, but baseline sample not adequately described
Study Attrition	Low	100% follow-up
Prognostic Factor Measurement	Moderate	Unclear validity, reliability of PF and unclear measurement properties of PF ("Afraid to start working again, because I don't think I will be able to manage")
Outcome Measurement	Low	Valid and reliable measurement of outcome (sick-listed Yes/No)
Study Confounding	High	Unclear variables included in stepwise models
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations
Study ID: Schultz 2004		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	High	Participation rate 43%; non-participants not adequately described
Study Attrition	Low	83% - 92% follow-up; no reasons for loss provided, but dropouts not significantly different from responders
Prognostic Factor Measurement	Low	Valid and reliable measure of PF (8 questions on RTW expectations)
Outcome Measurement	Low	Valid and reliable measurement of outcome (days lost due to low back disability in past 18 months, RTW status at 18 months)

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Study Confounding	Moderate	Minimal adjustment (LBP subgroup, SF-36 Health Transition, and Karasek's coworker support scale)
Statistical Analysis and Reporting	Moderate	No conceptual framework; data driven based on P values of univariate associations
Study ID: Shaw 2009		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	High	Participation rate unclear (volunteers); no description of eligible non-participants; full baseline sample not described
Study Attrition	Moderate	91% follow-up at 3 months; non-responders were different on age, gender, and education
Prognostic Factor Measurement	Moderate	Unclear validity and reliability of PF (full RTW at 4 weeks categorized, 'unlikely' grouped with 'not sure' for analyses)
Outcome Measurement	Moderate	Unclear validity and reliability measurement of RTW outcome (self-report RTW with unclear measurement properties); valid and reliable measurement of pain, function (NRS, RMDQ)
Study Confounding	Low	Adequate adjustment (age, gender, education, income, race, body mass index, smoking status, cause of injury, pain intensity rating, changes in pain since onset, missed a day of work already, job tenure, negative supervisor response, physical job demands, prior back surgery, past work absence due to LBP, employer allows modified duty, job enjoyment, worries about re-injury, frequency of moderate exercise, general health rating, felt downhearted and blue, felt under stress)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results
Study ID: Sherman 2009		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate 78%; non-participants not adequately described
Study Attrition	Low	91% follow-up 6 months and 1 year; reasons for loss provided and not likely to bias results
Prognostic Factor Measurement	Moderate	Unclear validity, reliability of PF measurement and unclear measurement properties of 1 of the 2 PFs (Likelihood of self-managing future back pain)
Outcome Measurement	Low	Valid and reliable measurement of outcome (RMDQ)
Study Confounding	Low	Adequate adjustment (baseline RMDQ score, baseline bothersomeness score, any disability, SF-36 mental health score, age, gender, education level, employment, medication use, duration of chronic LBP, pain travels below knee, days of LBP in last 3 months, intense LBP treatment, treatment group)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results; conceptual framework for inclusion of covariates

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Study ID: Steenstra 2005

Domain	Risk of bias level	Support for judgement
Study Participation	Low	Workers on sick leave required to submit baseline form to their employer (full participation)
Study Attrition	Low	100% follow-up
Prognostic Factor Measurement	Moderate	Unclear validity, reliability of PF measurement (sick leave > 10 or not)
Outcome Measurement	Low	Valid and reliable measurement of outcome (RTW, lasting RTW, total days sick leave)
Study Confounding	Moderate	Minimal adjustment (treatment by GP or specialist, seeking OP care, the interaction between the self-reported expected duration of sick leave and seeking OP care, complaints due to job stress, diminished mobility, and the interactions between expected duration of more than 10 days and seeking OP care and between seeking OP care and diminished mobility)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results; conceptual framework for inclusion of covariates

Study ID: Tran 2015

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate 52%; non-participants not adequately described
Study Attrition	Low	96% follow-up
Prognostic Factor Measurement	Moderate	Data-driven cut-point used to categorise a continuous measure of PF with good face validity (expectation of helpfulness of treatment; 0 - 10)
Outcome Measurement	Low	Valid and reliable measure of outcomes (NRS and RMDQ)
Study Confounding	Moderate	Minimal adjustment (age, sex, education, treatment arm, baseline SF-36 Physical Component Score, baseline RMDQ score); univariate data only available for meta-analyses
Statistical Analysis and Reporting	Moderate	Only mean and SD presented for change in each outcome by high and low expectations are presented; results of multivariate linear regression not presented due to non-significance of PFs

Study ID: Truchon 2012

Domain	Risk of bias level	Support for judgement
Study Participation	Moderate	Participation rate 77%; non-participants not adequately described
Study Attrition	Low	99% follow-up
Prognostic Factor Measurement	High	Unclear validity, reliability of PF measurement and unclear measurement properties of PF (Return to work expectations (time))

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Outcome Measurement	Low	Valid and reliable measurement of outcome (number of days of absence)
Study Confounding	Low	Adequate adjustment (work-related fear avoidance beliefs, annual family income, level of education, work schedule (irregular), work concerns)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results; conceptual framework for inclusion of covariates
Study ID: Turner 2008		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participation rate 51%; participants and non-participants adequately described; non-participants older and higher disability
Study Attrition	Low	100% follow-up
Prognostic Factor Measurement	Moderate	Continuous measure of PF with good face validity (NRS likelihood of recovery, 0 - 10) categorised for analyses without rationale
Outcome Measurement	Low	Valid and reliable measurement of outcome (wage replacement compensation for disability at 12 months)
Study Confounding	Low	Adequate adjustment (mental health, catastrophising, blame, relations with co-workers, work fear-avoidance, age, gender, race, education, pain intensity, RMDQ, "and other psychosocial variables")
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results; conceptual framework for inclusion of covariates
Study ID: Underwood 2007		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Low	Participation rate unclear; baseline sample and inclusion criteria adequately described
Study Attrition	Moderate	77% follow-up at 3 months, 75% follow-up at 1 year; no information provided on reasons or differences in characteristics
Prognostic Factor Measurement	Moderate	Unclear validity and reliability of PF measurement (3-pt Likert on treatment helpfulness)
Outcome Measurement	Low	Valid and reliable measurement of outcome (modified von Korff, RMDQ)
Study Confounding	Low	Adequate adjustment (manipulation and exercise, manipulation, exercise, additional education, working, age, male, pain and disability, quality of life, beliefs, episode length)
Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results; conceptual framework for inclusion of covariates
Study ID: Van Hooff 2014		
Domain	**Risk of bias level**	**Support for judgement**

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

	Low	Participation rate 84%
	Low	87% follow-up; authors indicate no differences between dropouts and those with follow-up data
	Low	Valid and reliable measure of PF (PSEQ)
	Low	Valid and reliable measure of outcome (ODI); clinical rationale provided for categorisation of continuous outcome
	Moderate	Minimal adjustment (function and employment); univariate only available for meta-analyses
	Moderate	Only mean and standard deviation of expectations measure presented for success and failure in outcome
Study ID: Van Wijk 2008		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	Participants from two trials (79% and 98% participation); selection criteria may lead to bias selected chronic group with no treatment response; response to facet joint block required
Study Attrition	Low	100% follow-up at 3 months, 82% at 1 year; reasons provided, no information on differences in characteristics
Prognostic Factor Measurement	High	Unclear which item of PCL is used for analyses, unclear description of PF
Outcome Measurement	Low	Valid and reliable measurement of outcome (VAS, Physical Activity Scale)
Study Confounding	Low	Adequate adjustment (baseline pain, 5 psychosocial domains: psychologically negative, adaptive manager, inflexible qualities, supporting partner, strong ego); no data available for meta-analyses (expectations reported as part of a factor)
Statistical Analysis and Reporting	Low	Analysis not sufficient for our study purposes only, expectations used to define psychological profile clusters for analyses
Study ID: Yelland 2006		
Domain	**Risk of bias level**	**Support for judgement**
Study Participation	Moderate	80% participation among eligible; inclusion criteria of high disability and failure to respond to conservative treatment may lead to bias
Study Attrition	Low	99% follow-up at 6 months, 96% at 1 year
Prognostic Factor Measurement	Low	Continuous measure of PF with good face validity (% improvement expected with treatment; 0 - 100)
Outcome Measurement	Low	Valid and reliable measurement of outcome (VAS)
Study Confounding	Moderate	Minimal adjustment (age, gender)

Table 8. Detailed QUIPS risk of bias assessments by study *(Continued)*

Statistical Analysis and Reporting	Low	Appropriate analysis for research question and study design; no apparent selective reporting of results

PF = prognostic factor; RMDQ = Roland Morris Disability Questionnaire; TEM = treatment effect modifier; RTW = return to work; VAS = pain visual analog scal; LBP = low back pain; PDI = Pain Disability Index; GCPS = Graded CHronic Pain Scale; NRS = pain numeric rating scale; ODI = Oswestry Disability Index; IPQ-R = Illness Perception Questionnaire (revised); CSQ = Coping Strategies Questionnaire; NASS = North American Spine Society Outcome Assessment scale; SF-36 = 36-item Short Forum Survey; BMI = body mass index; PQEQ = Pain Self-efficacy Questionnaire.

APPENDICES

Appendix 1. Search strategies for focused search using population ('back pain'), exposure ('expectations'), and study design ('prognosis') terms

Searches developed by Rachel Couban, Trials Search Coordinator, Cochrane Back Review Group; revised and executed by Leah Boulos, Evidence Synthesis Coordinator, Maritime SPOR SUPPORT Unit

Database: MEDLINE (Ovid)

1. dorsalgia.ti,ab.

2. exp Back Pain/

3. back pain.ti,ab.

4. backache.ti,ab.

5. back ache.ti,ab.

6. lumb* pain.ti,ab.

7. coccyx.ti,ab.

8. coccydynia.ti,ab.

9. sciatica.ti,ab.

10. exp sciatic neuropathy/

11. sciatic neuropathy.ti,ab.

12. spondylosis.ti,ab.

13. lumbago.ti,ab.

14. back disorder*.ti,ab.

15. back injur*.ti,ab.

16. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15

17. expectancy.tw.

18. expectation*.tw.

19. exp Attitude to Health/

20. Health Knowledge, Attitudes, Practice/

21. self efficacy/

22. self efficacy.tw.

23. illness belief*.tw.

24. ((disab* or self* or injur*) adj3 percept*).tw.

25. (outcome adj3 expect*).tw.

26. (questionnaire* adj3 (belief* or hope* or perceive* or expect* or desire* or percept* or likelihood or likely or anticipat* or want* or certainty or self-efficacy)).tw.

27. (recovery* adj3 (belief* or hope* or perceive* or expect* or desire* or percept* or likelehood or likely or anticipat* or want* or certainty or self-efficacy)).tw.

28. (measure* adj3 (belief* or hope* or perceive* or expect* or desire* or percept* or likelihood or likely or anticipat* or want* or certainty or self-efficacy)).tw.

29. 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28

30. Cohort Studies/

31. incidence.tw.

32. Mortality/

33. Follow-Up Studies/

34. prognos*.tw.

35. predict*.tw.

36. course.tw.

37. Survival Analysis/

38. 30 or 32 or 33 or 34 or 35 or 36 or 37 or 37

39. 16 and 29 and 38

Database: Embase (Embase.com)

1. dorsalgia:ti,ab

2. 'backache'/exp

3. 'back pain':ti,ab

4. backache:ti,ab

5. 'back ache':ti,ab

6. lumb* pain':ti,ab

7. coccyx:ti,ab

8. coccydynia:ti,ab

9. sciatica:ti,ab

10. 'sciatic neuropathy'/exp

11. 'sciatic neuropathy':ti,ab

12. spondylosis:ti,ab

13. lumbago:ti,ab

14. 'back disorder*':ti,ab

15. 'back injur*':ti,ab

16. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15

17. 'expectancy'/exp

18. expectancy

19. expectation*

20. 'attitude to health'/exp

21. 'attitude to disability'/exp

22. 'attitude to illness'/exp

23. 'self concept'/exp

24. 'self efficacy'

25. 'health belief'/exp

26. 'illness belief*'

27. (disab* OR self* OR injur*) NEAR/3 percept*

28. outcome NEAR/3 expect*

29. questionnaire* NEAR/3 (belief* OR hope* OR perceive* OR expect* OR desire* OR percept* OR likelihood OR likely OR anticipat* OR want* OR certainty OR 'self-efficacy')

30. recovery* NEAR/3 (belief* OR hope* OR perceive* OR expect* OR desire* OR percept* OR likelihood OR likely OR anticipat* OR want* OR certainty OR 'self-efficacy')

31. measure* NEAR/3 (belief* OR hope* OR perceive* OR expect* OR desire* OR percept* OR likelihood OR likely OR anticipat* OR want* OR certainty OR 'self-efficacy')

32. #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31

33. cohort analysis'/de

34. 'incidence'/de

35. 'mortality'/de

36. follow up'/de

37. 'survival'/de

38. 'prognosis'/de

39. prediction'/de

40. 'disease course'/de

41. #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40

42. #16 AND #32 AND #41

Database: CINAHL (with Full Text - EBSCOhost)

1. dorsalgia

2. (MH "Back Pain+")

3. back pain

4. backache

5. back ache

6. lumb* W1 pain

7. lumb* N5 pain

8. (MH "Coccyx")

9. (MH "Sciatica")

10. sciatica

11. coccyx

12. coccydynia

13. (MH "Lumbar Vertebrae")

14. lumb* N2 vertebra

15. (MH "Thoracic Vertebrae")

16. (MH "Spondylolisthesis") OR (MH "Spondylolysis")

17. lumbago

18. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17

19. expectancy

20. expectation*

21. (MH "Attitude to Health")

22. (MH "Health Beliefs")

23. (MH "Health Knowledge")

24. (MH "Self-Efficacy")

25. self efficacy

26. (MH "Attitude to Illness")

27. self perception

28. (MH "Self Concept")

29. S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28

30. "questionaire" OR (MH "Questionnaires+")

31. measurement

32. (MH "Recovery") OR "recovery"

33. S30 OR S31 OR S32

34. belief*

35. (MH "Hope") OR "hope"

36. desire*

37. likely

38. likelihood

39. (MH "Perception") OR "perception"

40. want*

41. anticipat*

42. certainty

43. S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42

44. S33 AND S43

45. S29 OR S44

46. (MH "Prospective Studies+")

47. (MH "Incidence")

48. (MH "Mortality")

49. follow up stud*

50. prognos*

51. predict*

52. (MH "Prognosis")

53. course

54. S46 OR S47 OR S48 OR S49 OR S50 OR S51 OR S52 OR S53

55. S18 AND S45 AND S54

Database: PsycINFO (EBSCOhost)

1. DE "Back Pain"

2. DE "Lumbar Spinal Cord"

3. back pain

4. DE "Spinal Column"

5. lumb* N2 vertebra*

6. coccyx

7. sciatica

8. lumbago

9. dorsalgia

10. back disorder*

11. DE "Back (Anatomy)"

12. (disc OR disk) N1 degenerat*

13. (disc OR disk) N1 herniat*

14. (disc OR disk) N1 prolapse*

15. failed back

16. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15

17. expectancy

18. expectation*

19. DE "Expectations"

20. DE "Health Attitudes"

21. DE "Health Knowledge"

22. DE "Health Behavior"

23. DE "Health Literacy"

24. DE "Self-Efficacy"

25. self efficacy

26. DE "Illness Behavior"

27. illness beliefs

28. (disab* OR self* OR injur*) N3 percept*

29. outcome* N3 expect*

30. questionnaire* N3 (belief* or hope* or perceive* or expect* or desire* or percept* or likelihood or likely or anticipat* or want* or certainty or self-efficacy)

31. recovery* N3 (belief* or hope* or perceive* or expect* or desire* or percept* or likelihood or likely or anticipat* or want* or certainty or self-efficacy)

32. measure* N3 (belief* or hope* or perceive* or expect* or desire* or percept* or likelihood or likely or anticipat* or want* or certainty or self-efficacy)

33. S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32

34. DE "Cohort Analysis"

35. incidence

36. DE "Mortality Rate"

37. DE "Morbidity"

38. DE "Followup Studies"

39. prognos*

40. predict*

41. course

42. DE "Prognosis"

43. survival analysis

44. S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43

45. S16 AND S33 AND S44

Appendix 2. MEDLINE and Embase search strategies for broad search using population (‘back pain’), and study design (‘prognosis’) terms (limiting potential reporting bias)

Screen of citations took advantage of existing work (searches to 2003 (Hayden 2007)) resulting in required screen load in this project of approximately 3500 citations. Searches developed by Rachel Couban, Trials Search Coordinator, Cochrane Back Review Group.

Database: MEDLINE (Ovid)

1 dorsalgia.ti,ab.

2 exp Back Pain/

3 backache.ti,ab.

4 exp Low Back Pain/

5 (lumbar adj pain).ti,ab.

6 coccyx.ti,ab.

7 coccydynia.ti,ab.

8 sciatica.ti,ab.

9 sciatic neuropathy/

10 spondylosis.ti,ab.

11 lumbago.ti,ab.

12 back disorder$.ti,ab.

13 or/1-12

14 Cohort Studies/

15 incidence.tw.

16 Mortality/

17 Follow-Up Studies/

18 prognos$.tw.

19 predict$.tw.

20 course.tw.

21 Survival Analysis/

22 or/14-21

23 13 and 22

Database: Embase

1 dorsalgia.mp.

2 back pain.mp.

3 exp LOW BACK PAIN/

4 exp BACKACHE/

5 (lumbar adj pain).mp.

6 coccyx.mp.

7 coccydynia.mp.

8 sciatica.mp.

9 exp ISCHIALGIA/

10 spondylosis.mp.

11 lumbago.mp.

12 back disorder$.ti,ab.

13 or/1-12

14 cohort analysis/

15 incidence/

16 mortality/

17 follow up/

18 survival/

19 prognosis/

20 prediction/

21 disease course/

22 or/14-21

23 13 and 22

Appendix 3. Modified QUIPS tool

Below we present a version of the QUIPS tool modified for this prognostic factor review. An electronic (MS Access) version of the full generic QUIPS tool is available at www.annals.org.

Summary: QUIPS identifies issues to consider for judging the overall risk of bias for a study. These issues will guide your thinking and judgement about the risk of bias within each of six domains. Some 'issues' may not be relevant to the specific study or the review research question. These issues are taken together to inform the overall judgement of potential bias for each of the six domains. Provide comments or text excerpts in the boxes below, as necessary, to facilitate the consensus process that will follow. Rate the adequacy of reporting for each applicable item as yes, partial, no or unsure, then (at the bottom of the page) rate potential risk of bias for each of the six domains as High, Moderate, or Low, considering all relevant issues.

BIAS: STUDY PARTICIPATION

Goal: To judge the risk of selection bias (likelihood that relationship between *PF* and *outcome* is different for participants and eligible non-participants).

Issues to consider for judging overall rating of risk of bias		Study methods and comments	Rating of reporting
Source of target population	The source population or population of interest is adequately described, including who the target population is (e.g. is the desired target population all workers? individuals filing compensation claims?), when (time period of study), where (location), and how (description of recruitment strategy). Comprehensive description would include characteristics of: individual (e.g. age, sex, depression), back pain (history of LBP, current functioning), work (type and characteristics of work environment), treatment (type and extent of care received) and social context (compensation status).		
Method used to identify population	The sampling frame and recruitment (e.g. newspaper advertisement, presentation to a health clinic, or captured from a claims database) are adequately described, including methods to identify the sample sufficient to limit potential bias (number and types used, e.g. referral patterns in health care).		
Recruitment period	Period of recruitment is adequately described.		
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described.		

(Continued)

Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described and should define a discreet group with LBP (e.g. the study may include physician diagnosis or explicit diagnostic codes).
Adequate study participation	There is adequate participation in the study by eligible individuals.
Baseline characteristics	The baseline study sample (i.e. individuals entering the study) is adequately described. Comprehensive description would include characteristics of: individual (e.g. age, sex, depression), back pain condition (history of LBP, current functioning), work (type and characteristics of work environment), treatment (type and extent of care received) and social context (compensation status).

Summary study participation:

The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between PF and outcome.

☐**Low risk of bias**

☐**Moderate risk of bias**

☐**High risk of bias**

BIAS: STUDY ATTRITION

Goal: To judge the risk of attrition bias (likelihood that relationship between PF and outcome are different for completing and non-completing participants).

Issues to consider for judging overall rating of risk of bias		**Study methods and comments**	**Rating of reporting**
Proportion of baseline sample available for analysis	Response rate (i.e. proportion of study sample completing the study and providing outcome data) is adequate.		
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described.		
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided.		
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for characteristics of: individual (e.g. age, sex, depression), back pain condition (history of LBP, current functioning), work (type and characteristics of work environment), treatment (type and extent of care received) and social context (compensation status).		

Summary study attrition:

Loss to follow-up (from baseline sample to study population analysed) is not associated with key characteristics (i.e. the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between individual recovery expectations and LBP outcome.

☐**Low risk of bias**

☐**Moderate risk of bias**

☐**High risk of bias**

BIAS: PROGNOSTIC FACTOR (PF) MEASUREMENT

Goal: To judge the risk of measurement bias related to how individual recovery expectations were measured (differential measurement of the prognostic factor related to the level of outcome).

Issues to consider for judging overall rating of risk of bias		Study methods and comments	Rating of reporting
Definition of the PF	A clear definition or description of individual recovery expectations is provided, capturing individual participant cognition (e.g. beliefs, perceptions, anticipations, expectations) and related to a future outcome. The description allows differentiation of general recovery expectations, treatment outcome expectations, and self-efficacy expectations.		
Valid and reliable measurement of PF	Method of 'individual recovery expectations' measurement is adequately valid and reliable to limit misclassification bias (e.g. may include relevant outside sources of information on measurement properties, also characteristics, such as limited reliance on recall). Examples of reliable and valid measurement of recovery expectations include the Revised Illness Perception Questionnaire to address recovery expectations (Timeline acute/chronic subscale) and treatment expectations (Treatment control subscale); and the Credibility/Expectancy Questionnaire.		
	Continuous variables are reported or appropriate cut-points (i.e. not data-dependent) are used.		
Method and setting of PF measurement	The method and setting of measurement of individual recovery expectations is the same for all study participants.		
Proportion of data on PF available for analysis	Adequate proportion of the study sample has complete data for the 'individual recovery expectations' variable.		
Method used for missing data	Appropriate methods of imputation are used for missing individual recovery expectations data.		

Summary prognostic factor measurement:

Individual recovery expectations are adequately measured in study participants to sufficiently limit potential bias.

☐**Low risk of bias**

☐**Moderate risk of bias**

☐**High risk of bias**

BIAS: OUTCOME MEASUREMENT

Goal: To judge the risk of bias related to the measurement of LBP outcome (differential measurement of outcome related to the baseline level of prognostic factor).

Issues to consider for judging overall rating of risk of bias		Study methods and comments	Rating of reporting
Definition of the outcome	A clear definition of the LBP outcome is provided, including duration of follow-up and ICF disability construct; return to work should be clearly defined if it means off work, work re-integration, work maintenance, or advancement.		
Valid and reliable measurement of outcome	The method of outcome measurement used is adequately valid and reliable to limit misclassification bias (e.g. may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and confirmation of outcome with valid and reliable test). Valid and reliable LBP outcome measures include: pain intensity, measured by a visual analogue scale (VAS) or other pain scale (e.g. numeric rating scale, or McGill pain score), functional limitations, measured by a LBP-specific scale (e.g. the Roland-Morris Disability Questionnaire, or the Oswestry Disability Index). Administrative return to work outcomes are considered valid.		
	Clear and appropriate cut-points for continuous outcome measures (i.e. not data-dependent) are used.		
Method and setting of outcome measurement	The method and setting of outcome measurement is the same for all study participants.		

Summary outcome measurement:

LBP disability outcome is adequately measured in study participants to sufficiently limit potential bias.

☐**Low risk of bias**

☐**Moderate risk of bias**

☐**High risk of bias**

BIAS: STUDY CONFOUNDING

Goal: To judge the risk of bias due to confounding (i.e. the effect of PF is distorted by another factor that is related to PF and outcome).

Issues to consider for judging overall rating of risk of bias		Study methods and comments	Rating of reporting
Important confounders measured	All important potential confounders are measured, including a reasonably comprehensive set of factors representing our domains of interest: individual demographics (e.g. age, sex, gender), social support (e.g., marital status, socioeconomic status), work factors and environment (e.g. occupation, physical demands, workplace culture), psychological factors (e.g. depression, anxiety, coping), and LBP complaint		

(Continued)

	factors (e.g. baseline pain severity, baseline disability, duration of episode at baseline).	
Definition of the confounding factor	Clear definitions of the important confounders measured are provided (e.g. including dose, level, and duration of exposures).	
Valid and reliable measurement of confounders	Measurement of all important confounders is adequately valid and reliable (e.g. may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall).	
Method and setting of confounding measurement	The method and setting of confounding measurement are the same for all study participants.	
Method used for missing data	Appropriate methods are used if imputation is used for missing confounder data.	
Appropriate accounting for confounding	Important potential confounders are accounted for in the study design (e.g. matching for key variables, stratification, or initial assembly of comparable groups; see variables below).	
	Important potential confounders are accounted for in the analysis (i.e. appropriate adjustment). Minimal control for potential confounding in included studies will consider 1 - 2 of the domains of interest. Adequate control for confounding will consider at least three of the five domains of interest. The domains of interest are: individual demographics (e.g. age, sex, gender), social support (e.g. marital status, socioeconomic status), work factors and environment (e.g. occupation, physical demands, workplace culture), psychological factors (e.g. depression, anxiety, coping), and LBP complaint factors (e.g. baseline pain severity, baseline disability, duration of episode at baseline).	

Summary study confounding:

Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between individual recovery expectations and LBP outcome.

☐**Low risk of bias**

☐**Moderate risk of bias**

☐**High risk of bias**

BIAS: STATISTICAL ANALYSIS and REPORTING

Goal: To judge the risk of bias related to the statistical analysis and presentation of results.

Issues to consider for judging overall rating of risk of bias	**Study methods and comments**	**Rating of reporting**

(Continued)

Presentation of analytical strategy	There is sufficient presentation of data to assess the adequacy of the analysis.
Model development strategy	The strategy for model building (i.e. inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model.
	The selected statistical model is adequate for the design of the study.
Reporting of results	There is no selective reporting of results.

Summary statistical analysis and reporting:

The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results, and selective reporting is unlikely.

☐**Low risk of bias**

☐**Moderate risk of bias**

☐**High risk of bias**

Appendix 4. Description of the six domains of the QUIPS

This description is paraphrased from Hayden 2013.

Study participation

The study participation domain addresses whether the study sample is representative of the population of interest. We will consider a study as having high risk of bias if the participation rate is low, a very selective rather than consecutive sample of eligible low back pain (LBP) individuals was recruited, or the study sample has a very different demographic and LBP characteristic distribution from our population of interest. Conversely, studies with high participation of eligible and consecutively-recruited LBP individuals who have characteristics similar to those in the source population would have low risk of bias.

Study attrition

The study attrition domain addresses whether participants completing the study (i.e. with follow-up data) represent the baseline sample. We will consider a study to have high risk of bias if it is likely that persons who completed the study differ from those lost to follow-up in a way that distorts the association between individual recovery expectations and LBP outcome. Conversely, studies with complete follow-up, or evidence that participants lost to follow-up are likely to be missing at random, will have low risk of bias.

Prognostic factor measurement

The prognostic factor measurement domain addresses adequacy of measurement of our factor of interest, individual recovery expectations toward non-differential measurement related to LBP disability. We will rate studies that use an unreliable method to measure individual recovery expectations or use different approaches for participants with different outcomes that may result in systematic misclassification as being at high risk of bias. Conversely, we will consider a study to have low risk of bias if individual recovery expectations are measured similarly (same method and setting) for all participants and use a valid, reliable measure, such as the Illness Perceptions Questionnaire.

Outcome measurement

The outcome measurement domain addresses the adequacy of LBP disability outcome measurement toward non-differential measurement related to recovery expectations. A study will have high risk of bias if there is likely to be differential measurement of outcome; for example, participants with negative expectations for recovery are assessed using a different approach from those with positive expectations. We will consider a study to have low risk of bias if the outcome is measured using the same method/setting for all participants and uses a valid, reliable measure (e.g. pain intensity by a visual analogue scale (VAS) or associated disability using the Roland-Morris Disability Questionnaire (RMDQ)).

Confounding

The study confounding domain addresses potential confounding, or distortion of the relationship between recovery expectations and LBP outcomes by another factor. A study will have high risk of bias if it does not control for any variables that have the potential to confound or explain the association between individual recovery expectations and outcome. Conversely, studies with adequate measurement of

important potential confounding variables and inclusion of these variables in a prespecified multivariable analysis will have low risk of bias. 'Minimal' control for potential confounding in included studies will consider some (one or two) of the confounding domains of interest, but do not control for a more robust set of confounders. We will judge 'adequate' control for confounding based on our proposed theoretical framework of the relationship between individual recovery expectations and LBP outcomes (Figure 1). This will include studies that adequately assess potential confounders, not on the proposed causal pathway, representing at least three of these domains: individual demographics (for example, age, sex, gender), social support (for example, marital status, socioeconomic status), work factors and environment (for example, occupation, physical demands, workplace culture), psychological factors (for example, depression, anxiety, coping), and LBP complaint factors (for example, baseline pain severity, baseline disability, duration of episode at baseline).

Statistical analysis and reporting

The statistical analysis and reporting domain addresses the appropriateness of the study's statistical analysis and completeness of reporting. We will consider a study to have low risk of bias if the statistical analysis is appropriate for the study design and data, if statistical model building is based on a conceptual framework or model (rather than a data-driven approach), and if all primary outcomes are reported.

Appendix 5. Guide to judge the quality of evidence for prognosis

Starting GRADE	Phase of investigation
HIGH	*Phase 3 Explanatory Study:* Explanatory research aimed at understanding prognostic pathways; or *Phase 2 Explanatory Study:* Explanatory research aimed at confirming independent associations between potential prognostic factor and the outcome
MODERATE	*Phase 1 Explanatory Study:* Explanatory research aimed at identifying associations between potential prognostic factors and the outcome, or Outcome prediction research providing evidence about prognostic factor associations

Downgrade if:		*Upgrade if:*	
Study limitations	Serious limitations when most evidence is from studies with moderate or unclear risk of bias for most bias domains	**Moderate or large effect**	For meta-analysis: pooled effect is moderate or large
	Very serious limitations when most evidence is from studies with high risk of bias for almost all bias domains		For narrative summary: moderate or large similar effect is reported by most studies
Inconsistency	Unexplained heterogeneity or variability in results across studies with differences in results not clinically meaningful. This may be supported by: - For meta-analysis: significant heterogeneity detected by test of heterogeneity and large I^2 value - For narrative summary: variations in effect estimates across studies with points of effect on either side of the line of no effect, and confidence intervals showing minimal overlap	**Exposure-gradient response**	For meta-analysis: gradient is present between analyses for factors measured at different doses For narrative summary: possible gradient exists within and between primary studies.
Indirectness	The study sample, the prognostic factor, and/or the outcome in the primary studies do not accurately reflect the review question		
Imprecision	For meta-analysis: (1) insufficient sample size and (2) no precise estimate of the effect size in the meta-analysis: confidence interval is excessively wide		

(Continued)

	and overlaps the value of no effect and contains values implying that the factor plays an important role in protecting or putting the individual at risk
	For narrative summary: Within-study imprecision, (1) sample size justification is not provided and there are fewer than 10 outcome events for each prognostic variable (for dichotomous outcomes) OR there are fewer than 100 cases reaching endpoint (for continuous outcomes); and (2) no precision in the estimation of the effect size within each primary study, AND
	Across-study imprecision: there are few studies and small numbers of participants across studies
Publication bias:	We recommend downgrading unless the value of the risk/protective factor in predicting the outcome has been repetitively investigated, ideally by phase 2 and 3 studies

Table modified (with permission) from Table 4, Huguet 2013

CONTRIBUTIONS OF AUTHORS

JAH is guarantor and conceived the review.

JAH, RR, RI, TP developed the protocol and JAH, MW, RO implemented the protocol.

MW, RO, JAH screened search results; MW and JAH carried out data extraction.

MW and JAH appraised the risks of bias and quality of the evidence.

MW and RO managed the data for the review and MW and JAH conducted analyses.

JAH and MW wrote the review report, with RO, RR, RI and TP providing methodological, clinical and general advice.

All authors reviewed the final version.

DECLARATIONS OF INTEREST

JAH has no known conflicts of interest related to the topic of this review. She is a Co-Convenor of the Cochrane Prognosis Methods Group and Advisory Board Member of Cochrane Back and Neck, however was not involved in editorial decisions involving this review. She has received peer-reviewed funding from the Canadian Institutes of Health Research, the Nova Scotia Health Research Foundation, and Cochrane. She previously held a Canadian Chiropractic Research Foundation/Dalhousie University Research Professorship in Epidemiology.

MW has no known conflicts of interest related to this review.

RDR has no known conflicts of interest related to the topic of this review. He receives payments for training courses provided in-house to other organisations; he receives funding from MRC and NIHR for other meta-analysis projects and has received payment from BMJ for review preparation. He is a Co-Convenor of the Cochrane Prognosis Methods Group, however was not involved in editorial decisions involving this review.

RI has no known conflicts of interest related to this review.

TP has no known conflicts of interest related to the topic of this review. She is employed full time by the Royal Holloway University of London and has received grants from Arthritis Research UK, and NIHR.

RO has no known conflicts of interest related to this review.

SOURCES OF SUPPORT

Internal sources

- Cochrane Methods Innovation Fund, UK.

External sources

- Canadian Institutes of Health Research, Canada.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

- We modified our theoretical framework of the relationship between recovery expectations and low back pain outcomes to provide more detailed guidance about important domains of characteristics to consider in 'Risk of bias' assessment (e.g. other potentially important covariates).
- We further operationalised the QUIPS 'Confounding' domain as follows, "We defined five domains of other covariates important for this review question: individual demographics (for example, age, sex, gender), social support (for example, marital status, socioeconomic status), work factors and environment (for example, occupation, physical demands, workplace culture), psychological factors (for example, depression, anxiety, coping), and low back pain complaint factors (for example, baseline pain severity, baseline disability, duration of episode at baseline). We defined 'minimally adjusted' study analyses as those presenting adjusted analyses controlling for 1-2 of these domains, and 'adequately adjusted' study analyses as those presenting adjusted analyses controlling for 3 or more of these domains".
- In the review we considered subacute and chronic populations together as subacute/chronic (≥ 6 weeks), whereas in the protocol we had planned to look separately at subacute (defined as 6 to 12 weeks) and chronic (> 12 weeks).
- We clarified inclusion criteria as intended in our study protocol: we excluded studies if they did not measure at least one of our primary outcomes, which include body function, activity limitation and participation restriction domains of the International Classification of Functioning, Disability and Health.
- For primary analyses, to balance homogeneity with availability of data, we used available study data from the time period closest to 12 months (defined as 'long, closest to 12 months'); this was not predefined in the protocol.
- We defined expectations reference time periods for subgroup analysis as: 1-month, 6-month, none or unclear reference period; this was not predefined in the protocol.
- If available in sufficient numbers, we had planned to separately extract and analyse continuous outcomes on a continuous scale, and hazard ratios for studies providing this measure of association. This was not possible.
- For consistency, we recalculated associations to be in the same direction, as necessary, with odds ratios above 1 indicating that better (positive) expectations are associated with a better (improved) outcome.
- As defined in our protocol, but not clearly, we included an additional primary outcome, 'important recovery', that was available in studies as a dichotomous measure of clinically important recovery in functional limitations, pain intensity, and/or work participation.
- We included an additional follow-up period, 'very long' (> 16 months) that was not predefined in the protocol.
- We did not conduct sensitivity analyses to explore the robustness of results excluding studies with mixed pain or specific low back pain populations as we had planned.
- We added a note about the presentation of forest plots and MA results in text, "We present forest plots of meta-analyses in the text of this review when three or more studies were available for meta-analyses in primary analyses, and when at least three studies were available for two or more subgroups in subgroup analyses."
- There were not enough studies available to allow other planned sensitivity analyses for studies including only low back pain populations versus studies including a small proportion of mixed pain populations, surgical candidates or individuals with lumbar disc herniation.
- Due to the format of our data, we used Egger's test for potential publication bias, although our protocol had described a plan to use Peter's test for dichotomous expectations measure and work participation outcome.This may contribute to a false-positive result.
- Our protocol included description of related methods projects, including investigation of the impact of various search strategies, and refinement and further guidance of the QUIPS tool. This work is ongoing and will be reported separately.