Early Psychological Interventions for Somatic Symptom Disorder and Functional Somatic Syndromes: A Systematic Review and Meta-Analysis.

Online Supplement

Lukas Berezowski

University of Hamburg

Lea Ludwig

University of Hamburg

Alexandra Martin

University of Wuppertal

Bernd Löwe

University Medical Center Hamburg-Eppendorf

Meike C. Shedden-Mora

University Medical Center Hamburg-Eppendorf

Contents

Online Supplement A

Protocol

REVIEW PROTOCOL

Early psychological interventions for somatic symptom disorders and functional somatic syndromes: Protocol for a systematic review and meta-analysis.

Meike Shedden Mora[1,], Lukas Berezowski[2], Lea Ludwig[2], Alexandra Martin[3], & Bernd Löwe[1]*

Author Affiliations:
[1] Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20246 Hamburg, Germany
[2] Clinical Psychology and Psychotherapy, Institute of Psychology, Faculty of Psychology and Movement Sciences, University of Hamburg, Von-Melle-Park 5, 22146 Hamburg, Germany
[3] Department of Clinical Psychology and Psychotherapy, University of Wuppertal, Max-Horkheimer-Straße 20, 42097 Wuppertal, Germany

E-Mails:
Meike Shedden Mora*: m.shedden-mora@uke.de
Lukas Berezowski: lukas.berezowski@studium.uni-hamburg.de
Lea Ludwig: lea.ludwig@uni-hamburg.de
Alexandra Martin: martin@uni-wuppertal.de
Bernd Löwe: b.loewe@uke.de

*Corresponding author

Contributions
MSM, BL and AM developed the research question and conceptual background for this review. MSM, LL and LB developed the outline for this review. LB, MSM and LL formulated the search terms. LB and MSM will perform the literature search, data extraction and data analysis. BL, AM and LL will provide regular feedback on the progress and results. MSM, LB, BL, LL and AM will prepare the manuscript for publication and will be responsible for its content.

Amendments
Amendments will be approved in consensus between all authors.
If we decide to amend the protocol, we will provide the date of amendment, name the section in which the amendment occurs, explicate the amendment and explain its rationale. In order to ensure transparency and comprehensibility, amendments will be documented separately. LB is responsible for documenting and implementing amendments.

Sources
Internal funding

Sponsor
Internal funding. The study is supported by the European Research Network for Persistent Somatic Symptoms (EURONET-SOMA), but does not receive funding.

Role of sponsor
N.A.

Guidelines and registration
This protocol is developed in accordance with the PRISMA-P guideline (Shamseer et al., 2015). This review will be registered with the International Prospective Register of Systematic Reviews (PROSPERO).

Keywords
Persistent somatic symptoms; functional somatic syndrome; somatic symptom disorder; bodily distress; somatoform disorder; early psychological intervention; prevention.

Introduction

Rationale
Somatic symptom disorders (SSD) and functional somatic syndromes (FSS) pose a major challenge for health care. The term SSD has been introduced in the current DSM-5 diagnostic classification and replaces the former DSM-IV diagnoses of somatization disorder, undifferentiated somatoform disorder and pain disorder. The term FSS refers to symptoms that can typically be attributed to one organ system but do not correlate to a well-defined structural organic pathology (Henningsen, Zipfel, Sattel, & Creed, 2018). In the following, these two terms will be used to describe burdensome persistent physical symptoms that are present for at least several months (Henningsen, Gündel et al., 2018). It is relevant to note, that other terms, such as bodily distress syndrome/disorder or medically unexplained (physical) symptoms are also used with slightly different connotations but considerable overlap in diagnostic features.
SSD/FSS show 12-month prevalence rates of 5% up to 16% among the European population (Petersen, 2019; Wittchen et al., 2011). At the severe end of the continuum from mild to disabling bodily complaint, SSD/FSS cause substantial suffering, go along with comorbid depression and anxiety, reduced quality of life, and lead to high disability and high health care costs (Henningsen, Zipfel, et al., 2018; Konnopka et al., 2012; Löwe et al., 2008). SSD/FSS are under-recognized and their detection is often limited to very severe cases (Schaefert et al., 2010).
While psychological therapies are currently the most effective treatment option for SSD/FSS, effect sizes are generally only small to moderate (Abbass, Kisely, & Kroenke, 2009; Hausteiner-Wiehle et al., 2012; Henningsen, Zipfel et al., 2018; Kleinstäuber, Witthöft, & Hiller, 2011; Koelen et al., 2014; Kroenke, 2007; Van Dessel et al., 2014; van Gils et al., 2016). One possible explanation for the small effect sizes could be the high level of chronicity in these patients. The mean symptom durations reported in current reviews and meta-analyses evaluating psychological interventions for SSD/FSS revealed – if reported in the studies – symptom durations ranging from 3 to 25 years (Kleinstäuber et al., 2011; Koelen et al., 2014; Van Dessel et al., 2014; van Gils et al., 2016).
Attempts to detect and treat patients with high somatic symptom burden in primary care as early as possible, such as our Sofu-Net study (Löwe et al., 2017; Shedden-Mora et al., 2016), are promising and have been successful in improving rates of patients receiving mental health care. However, the estimated mean duration of untreated illness in our primary care sample of patients with somatoform disorders was 25 years (Herzog, Shedden-Mora, Jordan, & Löwe, 2018). Similarly, other studies such as the PROSPECTS study have reported mean symptom durations of 10.5 years in patients with persistent physical symptoms from primary, secondary and tertiary care (Claassen-van Dessel, van der Wouden, Hoekstra, Dekker, & van der Horst, 2018). These durations of untreated symptoms by far

exceed those of other mental disorders such as depression (Kisely, Scott, Denney, & Simon, 2006; Okuda et al., 2010). Thus, chronicity might well partly explain the small effect sizes achieved by psychological treatments. Aiming at detecting and treating patients with SSD/FSS as early as possible therefore seems a promising approach to improve treatment outcome and prevent the chronic long-term course and related suffering for these patients.

Currently, there is no systematic evidence of the effectiveness of specific early intervention approaches for SSD/FSS. Several studies have tried to target somatic symptoms in early interventions for specific functional somatic syndromes. These include psychological interventions for subacute lower back pain (del Pozo-Cruz et al., 2012), whiplash injuries (Brison et al., 2005; Oliveira, Gevirtz, & Hubbard, 2006), and temporomandibular disorder-related pain (Gatchel, Stowell, Wildenstein, Riggs, & Ellis, 2006). For patients with FSS or high somatic symptom burden in general, two studies on primary care physician-delivered enhanced care targeted patients presenting with a new health problem, which included a large proportion of subacute patients (Rosendal et al., 2007; Toft et al., 2010). In this non-systematic way, evidence suggests that early psychological interventions might be effective in reducing symptoms, reducing the risk of developing a chronic timeline, improving illness consequences, and reducing costs. However, evidence needs to be established systematically.

To the best of our knowledge, this is the first systematic review on early intervention approaches for somatic symptom and related disorders, functional somatic syndromes, somatoform disorders, medically unexplained (physical) symptoms, and bodily distress syndromes. If effective, early interventions provide a more efficient way to manage patients with SSD/FSS and prevent the chronic development of symptoms. Early interventions will gain increasing importance and could be implemented in routine health care.

Objective

The aim of this systematic review and meta-analysis is to systemically examine the efficacy of early psychological interventions in preventing and treating SSD/FSS compared to control treatments in adults.

Methods

The methods of this systematic review and meta-analysis were developed by consulting the PRISMA-P guideline (Shamseer et al., 2015), the Cochrane Handbook for Systematic Reviews of Interventions (Higgins & Green, 2011) and further literature on conducting systematic reviews and meta-analyses (e.g. Borenstein, Hedges, Higgins, & Rothstein, 2009).

Eligibility criteria

Participants:

Participants need to be adult humans (18 years and older) fulfilling at least one of the following criteria:

1.) being at elevated risk for developing a SSD/FSS due to an acute event (e.g. whiplash trauma after car accident, infection, surgery) *(prevention population, 'incident' definition)*
2.) suffering from a SSD/FSS as diagnosed by a medical/mental health professional for a maximum of 12 months, or suffering from sub-threshold functional symptoms, or exhibiting somatic symptoms without clear somatic etiology and indication for somatic treatment *(early intervention population, 'time' or 'recent onset' definition)*
3.) first presentation with an SSD/FSS to health care provider (*first presentation population, 'help-seeking' definition)*

The onset of SSD/FSS cannot clearly be defined in many cases. Therefore, we incorporated these three different participant eligibility criteria to capture the whole population of interest for this review. The first criterion refers to populations with a known elevated likelihood of developing an SSD/FSS due to a specific event. Possible risk events are e.g. motor vehicle accidents leading to whiplash injuries (Barnsley, Lord, & Bogduk, 1994), suffering from an acute gastroenteritis (Löwe et al., 2016; Thabane & Marshall, 2009) or surgery (Bruce & Quinlan, 2011). In this at-risk population, interventions might effectively target SSD/FSS, preventively (e.g. Oliveira et al., 2006). The second criterion refers to populations with sub-threshold symptoms of SSD/FSS, often framed as "medically unexplained symptoms" (e.g. Nimnuan, Hotopf, & Wessely, 2001), or new onsets of full-blown SSD/FSS. The third criterion refers to populations who seek professional help for their full-blown SSD/FSS for the first time, irrespective of the duration of the disorder. Here, the earliness of the intervention is defined via the help-seeking behavior of the affected population, in contrast to the duration of illness incorporated in the second criterion. By investigating the efficacy of early psychological interventions in this population, we aim at resembling more closely the conditions in routine care where the delivery of early psychological interventions is possible when the affected individuals do seek help in the first place, only.

Study data will be included when the whole sample of a study fulfills the criteria or when data for these participants are reported separately. We restrict our review to an adult population since the impetus for our research question originates from research on adults and interventions appropriate for children and adolescents might differ substantially from interventions appropriate for adults.

Interventions:
We will include studies evaluating the efficacy of early psychological interventions in preventing or treating SSD/FSS. We define psychological interventions as treatments intending to induce change in behavior, emotion and/or cognition via psychological means. When addressed at a population fulfilling our participant inclusion criteria, we conceive psychological interventions as early psychological interventions.
Studies addressing clinician-directed interventions will be included, when these interventions aim at fostering the use of psychological interventions in clinicians and patient-level outcomes are reported. In this case, patients still need to fulfill the participant criteria mentioned above.
In accordance with our definition of psychological interventions, studies examining the efficacy of pharmacological or physiotherapeutic interventions will be excluded.

Comparators:
The early psychological intervention must be compared to no treatment, standard medical care or treatment as usual, wait-list control group or placebo group.

Outcomes:
Studies do not need to report specific outcomes in order to be included in the narrative review. For inclusion in the meta-analyses, however, studies need to report at least one of the following either primary or secondary outcomes at post-treatment or at follow-up measurement:

1.) primary outcomes:
    - somatic symptom severity (self-report)
    - health-related quality of life (self-report)
2.) secondary outcomes:

- unwanted negative treatment effects
- diagnostic status concerning SSD/FSS (clinician-rated)
- anxiety
- depression
- health care utilization (i.e., number of doctor visits)
- consumer satisfaction (self-report)

Outcomes were selected based on recommendations for research on interventions for SSD/FSS (Rief et al., 2017).

Study designs:
We will include prospective randomized-controlled trials, including cluster randomized trials.

Language:
We will include studies reported in English or German.

Years considered:
We will look for study data published from 1st January 1994 until 1st September 2019. The year 1994 was chosen to include studies after the introduction of DSM-IV and ICD-10.

Filters:
Irrelevant studies will be filtered out during literature search by using filters for randomized controlled trials and for publications after 1994 (see Search strategy).

Information sources
The following databases (and their providers) will be used to obtain data:

- PubMed (NCBI)
- PsycINFO (Ovid)
- Web of Science (Clarivate Analytics)

We decided to search PubMed and PsycINFO in addition to Web of Science since our review question addresses a topic of both medical and psychological interest.
Further relevant studies will be searched by conducting a backward search using the included studies. For the backward search, reference lists of included studies will be scanned for further potentially relevant studies.
MSM and LB developed and will carry out the search.

Search strategy
To our knowledge, no past review has aimed at covering the full range of SSD/FSS. Thus, we developed a more comprehensive search strategy based on previously published studies, reviews, textbooks and our expertise in order to cover as many clinical conditions and diagnoses as possible (see description of search strategy in the appendix).
The search strategy was developed by MSM, LL and LB. Using potentially eligible articles in the authors' bibliographies, LB piloted and refined the search strategy.
The search strategy for the electronic databases is best described as a conjunction of two parts. The first part consists of search terms for SSD/FSS, while the second part consists of search terms narrowing the results on studies investigating early psychological interventions. Relevant studies will be searched

separately for each clinical condition. Thus, while the 1st part of the search strategy varies between searches, the 2nd part remains constant.
The search will be limited to titles and abstracts of articles and phrase searching will be used for compound search terms in order to reduce irrelevant search results. If available, filters incorporated in the electronic databases limiting the search to randomized-controlled trials and studies published from 1st January 1994 until 1st September 2019 will be used.
As an example, the search for studies examining early psychological interventions for irritable bowel syndrome in PubMed will be as follows:

("Irritable bowel"[Title/Abstract] OR "Irritable colon"[Title/Abstract] OR IRS[Title/Abstract] OR "Mucous colitis" [Title/Abstract] OR "Mucous colitides" [Title/Abstract])
AND
("Early intervention" [Title/Abstract] OR "Early interventions" [Title/Abstract] OR "Early therapy" [Title/Abstract] OR "Early therapeutic" [Title/Abstract] OR "Early treatment" [Title/Abstract] OR "Early treatments" [Title/Abstract] OR "Early management" [Title/Abstract] OR "Early psychotherapy" [Title/Abstract] OR "Early psychotherapeutic" [Title/Abstract] OR "Early CBT" [Title/Abstract] OR "Early psychoeducation" [Title/Abstract] OR "Early psychoeducational" [Title/Abstract] OR "Early psycho-education" [Title/Abstract] OR "Early psycho-educational" [Title/Abstract] OR "Early education" [Title/Abstract] OR "Early educational" [Title/Abstract] OR "Early self-help" [Title/Abstract] OR "Early self help" [Title/Abstract] OR "Early information" [Title/Abstract] OR "Early rehabilitation" [Title/Abstract] OR "Early bibliotherapy" [Title/Abstract] OR "Early bibliotherapeutic"[Title/Abstract] OR (("new onset" [Title/Abstract] OR "recent onset" [Title/Abstract] OR sub-acute[Title/Abstract] OR acute[Title/Abstract] OR sub-threshold[Title/Abstract] OR sub-clinical[Title/Abstract] OR non-chronic[Title/Abstract]) AND (intervention[Title/Abstract] OR interventions[Title/Abstract] OR therapy[Title/Abstract] OR treatment[Title/Abstract] OR treatments[Title/Abstract] OR management[Title/Abstract] OR psychotherapy[Title/Abstract] OR CBT[Title/Abstract] OR psychoeducation*[Title/Abstract] OR psycho-education*[Title/Abstract] OR education*[Title/Abstract] OR self-help[Title/Abstract] OR "self help" [Title/Abstract] OR information[Title/Abstract] OR rehabilitation[Title/Abstract] OR bibliotherap*[Title/Abstract])) OR prevent[Title/Abstract] OR preventary[Title/Abstract] OR preventive[Title/Abstract] OR preventative[Title/Abstract] OR preventing[Title/Abstract] OR prevention[Title/Abstract] OR "psychological first aid"[Title/Abstract])

We will not search systematically for grey literature, e.g. dissertations, theses or presentations. The full search strategy including covered clinical conditions and search terms is described in the appendix.

Study records

Data management

For each electronic database search, we will document the number of identified records. Results of all searches will be exported to EndNote (Version X9.2) and deduplicated using the built-in deduplication function.

Selection process

The selection process is composed of two phases.
In the first phase, titles and abstracts of search results from electronic databases will be screened by MSM and LB independently against the eligibility criteria. Studies which seem to fulfill eligibility criteria

or where eligibility is uncertain will proceed to full-text screening. At the stage of full-text screening, a subset of 30 studies will be screened by MSM and LB independently and in duplicate in order to establish inter-rater agreement. The further selection process will be selected depending on the level of inter-rater agreement, i.e., independent screening by LB, or independent and duplicate screening by MSM and LB. In the latter case, disagreements will be resolved by discussion. When disagreements cannot be resolved by discussion, LL will be asked to arbitrate. When full texts cannot be accessed, we will locate and contact the corresponding study authors via email to obtain the full text with a second attempt when we receive no response within two weeks. Reasons for exclusion will be documented according to the following prioritization: no prospective randomized-controlled design, study sample does not fulfil eligibility criteria, no psychological intervention, no adequate comparator group, publication beyond time frame of interest, other language than English or German.
The second phase consists of the backward search. For this purpose, reference lists of included studies will be screened for further potentially eligible studies following the above-mentioned procedure. Studies which seem to fulfill eligibility criteria or where eligibility is uncertain will be checked for prior inclusion or exclusion decision during the first phase. Studies which have not been included or excluded in prior steps of the selection process, will proceed to full-text screening. The procedure for full-text screening is identical to the procedure in the first phase. The backward search will be repeated until no further potentially eligible studies are detected.

Authors will not be blinded to any aspect of identified studies during the study selection process.

Data collection process
After finishing the selection process, authors will discuss whether they noticed any signs of duplicate reports. If so, we will look for cross-references and compare authorship, sample characteristics and outcome characteristics (von Elm, Poglia, Walder, & Tramèr, 2004). Data from duplicate reports will be treated as stemming from one study. For our analyses, we will use data from the original report defined by being the oldest and/or largest one. If data of interest is not available in the original report, we will use data from duplicate reports.
Data will be collected using a standardized form implemented in Microsoft Access 2016. The standardized form will be developed by LB and reviewed by MSM and LL. A subset of 10 studies will be coded by MSM and LB in duplicate in order to establish inter-rater agreement in outcome data. Data collection will be conducted unblinded. When information necessary for effect size calculation is missing in a report, we will locate and contact the corresponding study authors via email to obtain further information, with a second attempt when we receive no response within two weeks.

Data items
We will extract the following data from primary studies:
General information:
- authors
- publication year
- corresponding author email address
- report language
- country where study was conducted
- type of study design (RCT vs. cluster-RCT)

Participants:
- total sample size

- type of participant population (prevention, early intervention, first presentation or a combination of these)
- disorder/syndrome of interest (for prevention: at risk; for early intervention and first presentation: present)
- eligibility criteria
- mean age
- SD age
- proportion female
- mean duration of symptoms or disorder

Intervention:

- type of intervention (e.g. psychoeducation, CBT, psychodynamic therapy)
- type of delivery (e.g. face-to-face, web-based, written material)
- person delivering the intervention (e.g. nurse-led, physician-led)
- intervention intensity (low: no/one contact with professional; high: repeated contact with professional)
- target of intervention (patient-centered vs. clinician centered)
- number of treatment sessions
- type of control group

For each outcome of interest:

- type of outcome (somatic symptom severity, anxiety etc.)
- measure
- source (self-report vs. clinician-rated)
- higher value in outcome measure desirable (yes vs. no)
- time point of measurement (with end of treatment = 0)
- for continuous outcomes: means, standard deviations and sample sizes (or other data to calculate effect sizes)
- for dichotomous outcomes: number of (non-)events in each group, sample sizes (or other data to calculate effect sizes)

For cluster-randomized trials, additionally:

- statistical analysis accounting for clustering (yes vs. no)
- numbers of clusters in intervention and in control group
- mean cluster size
- intracluster correlation coefficient

For each study, we will extract outcome data at three time-points: baseline, end of treatment as well as longest follow-up measurement. When end of treatment measurement was not reasonable to conduct e.g. due to the shortness of the studied intervention (e.g. one psychoeducation session), we will extract data from the first measurement after end of treatment.

When several early psychological interventions are delivered in different treatment arms, we will collapse data of these treatment arms (Borenstein et al., 2009). When multiple control groups fulfilling the eligibility criteria are reported, we will extract data from the most active control treatment (e.g. placebo group > TAU).

When a study reports multiple effects for an outcome, e.g. due to employing multiple measures for the same construct, we will extract data from the main outcome measure of a given study. When study authors did not select a main outcome measure, we will extract data from the most valid and reliable measure.

When means and standard deviations are not sufficiently reported to calculate effect sizes for continuous outcomes, we will extract other statistics to calculate effect sizes or rely on reported effect sizes, alternatively (see Borenstein, 2009). If effect size calculation is not possible anyway, we will contact study authors as described above (see *Data collection process*). The same procedure will be applied for dichotomous outcomes, when events per condition and the respective sample sizes are not sufficiently reported. Whenever possible, we will use results from intention-to-treat analyses when evaluating effect size. When necessary and appropriate, we will convert between effect size metrics to obtain the desired one (Borenstein et al., 2009).

Outcomes and prioritization

Primary outcomes

Primary outcomes will be somatic symptom severity and health-related quality of life.

Somatic symptom severity subsumes all self-report measures of somatic symptoms related to the studied SSD/FSS or more generally somatization. Examples for measures are numeric ratings scales for pain, the BDS checklist (Budtz-Lilly et al., 2015), or the Patient Health Questionnaire-15 (Kroenke, Spitzer, & Williams, 2002). Since symptom patterns differ between the clinical conditions of interest in this review, we will integrate measures of different types of symptoms (e.g. pain, fatigue) between studies.

As a second primary outcome, we included health-related quality of life. We define health-related quality of life as the individual's perceived health status covering factors like functioning, disability and well-being (Karimi & Brazier, 2016; Moons, 2004). We will include data from self-report measures of health-related quality of life, e.g. the SF-36 (Ware, 2000). We chose health-related quality of life as second primary outcome since health-related quality of life seems to be an important outcome from patient perspective and is not solely determined by the presence or severity of symptoms (Smith, Avis, & Assmann, 1999; Spiegel et al., 2004; Testa & Simonson, 1996). Furthermore, treatment recommendations for SSD/FSS conceive restoring functioning and learning to cope with symptoms as important treatment goals (Henningsen, Zipfel et al., 2018; Roenneberg, Hausteiner-Wiehle, Schäfert, Sattel, & Henningsen, 2018).

Secondary outcomes

Secondary outcomes will comprise of unwanted negative treatment effects, diagnostic status concerning functional condition, anxiety, depression, health care utilization (doctor visits), and consumer satisfaction.

We decided to include unwanted negative treatment effects as second primary outcome to enable a balanced evaluation of early psychological interventions. Since data on unwanted negative treatment effects in psychological interventions are scarce (Rief et al., 2017) and important unwanted negative effects in the treatment of SSD/FSS seem unclear, we make no further specifications.

Diagnostic status is a dichotomous outcome defined as the presence of an SSD/FSS as established by a clinician via a valid method (e.g. structured interview, medical examination). Although this outcome is closely related to somatic symptom severity, we added this outcome due to its significance for clinical practice and decision-making.

We decided to include anxiety and depression as secondary outcomes since they represent conditions frequently comorbid with SSD/FSS and are associated with outcome and functioning (Creed et al., 2005; De Waal, Arnold, Eekhof, & Van Hemert,2004; Henningsen, Zimmermann, & Sattel, 2003). We will include both self-report and clinician-rated measures.

Health care utilization is operationalized as doctor visits, describing the frequency of participants seeking outpatient treatment. We will include data on doctor visits if quantified via either objective data (e.g. medical records) or self-report. We decided to include this outcome in order to reflect the potential health-economic effect of early psychological interventions. However, this outcome does not represent the cost-effectiveness of early psychological interventions, since we do not consider the costs of implementing early psychological interventions in our analysis.
Consumer satisfaction reflects the acceptance of the treatment as reported by the participants. We included this measure since consumer satisfaction should be an important criterion when considering the implementation of an intervention in routine health care.

Risk of bias of individual studies
Risk of bias will be assessed using the Revised Cochrane risk-of-bias tool for randomized trials (RoB 2; Higgins, Savović, Page, & Sterne, 2019). The RoB 2 assesses biases using multiple items for each of the following domains: bias arising from the randomization process, bias due to deviations from intended interventions, bias due to missing outcome data, bias in measurement of the outcome and bias in selection of the reported result. The assessment procedure results in a judgement of bias for each domain with the categories low risk, some concerns and high risk.
Risk of bias assessment will be conducted by LB. LB will be unblinded to all study information during assessment. Decisions will be checked by MSM. Disagreements will be resolved by finding a consensus in discussion. If consensus cannot be established, LL will be asked to arbitrate.
As recommended in the guidance document of the tool (Higgins et al., 2019) we will report domain-level judgements of risk of bias narratively and graphically to inform evaluation of treatment efficacy. We do not intend to incorporate risk of bias ratings in our statistical analyses.

Data synthesis
Criteria for conducting a meta-analysis
Data of study characteristics will be described narratively and descriptively. Meta-analyses will be performed when at least three studies are available for the respective analysis. If a meta-analysis is not appropriate, we will report study outcomes narratively (see *Narrative synthesis*).

Planned analyses
Meta-analyses will be conducted within the statistical software R (Version 3.6; R Core Team, 2019). For each outcome at each time point (post-treatment vs. follow-up) we will conduct random-effects analyses since we do not conceive the collected effect sizes to represent a single population effect size due to heterogeneity, e.g. in interventions, populations and outcome measures. Weights will be computed using the inverse-variance method. Between-study variance (τ^2) will be estimated using the method of restricted maximum likelihood according to the recommendations by Langan et al. (2018). We will report I^2 together with its respective 95% confidence interval to ease interpretation of the heterogeneity estimate. For each outcome, we will report a summary effect and its corresponding 95% confidence interval using the Knapp-Hartung method (Inthout, Ioannidis, & Borm, 2014; Knapp & Hartung, 2003) as well as its 95% prediction interval (Inthout, Ioannidis, Rovers, & Goeman, 2016).
For diagnostic status data, we will compute and report risk ratios, with numbers < 1 representing more desirable results in the intervention group. If a study included in the analysis reports 0 events in a cell, we will add 0.5 to all cells of the respective matrix to allow calculation of risk ratios. For all other outcomes, we will compute and report Hedge's *g* (Hedges, 1981), with positive numbers representing more desirable effects in the intervention group.

When outcome data are differentially poled between studies, data will be adjusted before analysis so that all scales are aligned. If data from cluster-randomized trials were not analyzed accounting for clustering effects in a given study, we will approximate correct data by inflating standard errors as described by Higgins, Deeks & Altman (2011). Analogous to Van Dessel et al. (2014), we will impute an intracluster correlation of 0.031 based on an estimation from Campbell, Fayers & Grimshaw (2005), when information on intracluster correlation is missing,

Additional analyses

We will use meta-regression to analyze the impact of moderators on the treatment effect size. As moderators, we will examine the intensity of interventions, duration of symptoms, type of participant population as well as type of control group.

Furthermore, we will investigate whether effects at follow-up vary as a function of length of follow-up using meta-regression for each outcome.

Finally, we will examine the relationship of all moderators included in our additional analyses with other study-level variables, descriptively, in order to detect potential confounding.

Sensitivity analyses

We will investigate the robustness of the results to the method employed for estimating between-study variance. For this purpose, we will repeat analyses with τ^2 estimated via the two-step DerSimonian-Laird method (DerSimonian & Kacker, 2007) since it is recommended as an alternative to the restricted maximum likelihood method (Langan et al., 2018).

Additionally, we will conduct a meta-regression with type of study design as binary predictor to test whether results differ when calculating separate effects for randomized-controlled trials and cluster-randomized trials.

Narrative synthesis

When meta-analysis is not appropriate, we will describe the included population, the employed intervention and its effect for each study, narratively.

Meta-bias(es)

As recommended by Carter, Schönbrodt, Gervais, & Hilgard (2019) we will explore the range of possible outcomes when correcting for meta-biases by implementing multiple methods. We decided to implement the conditional PET-PEESE procedure (Stanley & Doucouliagos, 2013) as well as the 3PSM procedure (Vevea & Hedges, 1995) since both methods have been recommended by Carter et al. (2019). Moreover, these procedures seem to be appropriate for the expected conditions of our meta-analyses according to the simulation data provided by Carter et al. (2019) (severity of publication bias = high, heterogeneity = 0.4, number of studies = 10, questionable research practice environment = medium, true effect size = 0 or 0.5).

Confidence in cumulative estimate

We will use the GRADE approach (Schünemann, Brożek, Guyatt, & Oxmann, 2013) to assess the confidence in cumulative estimate.

References

Abbass, A., Kisely, S., & Kroenke, K. (2009). Short-term psychodynamic psychotherapy for somatic disorders. Systematic review and meta-analysis of clinical trials. *Psychotherapy and Psychosomatics, 78*(5), 265-274. doi:10.1159/000228247

Barnsley, L., Lord, S., & Bogduk, N. (1994). Whiplash injury. *Pain, 58*(3), 283-307. doi:10.1016/0304-3959(94)90123-6

Borenstein, M. (2009). Effect sizes for continous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis (2nd edition).* New York: Russel Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester: John Wiley & Sons, Ltd.

Brison, R. J., Hartling, L., Dostaler, S., Leger, A., Rowe, B. H., Stiell, I., & Pickett, W. (2005). A randomized controlled trial of an educational intervention to prevent the chronic pain of whiplash associated disorders following rear-end motor vehicle collisions. *Spine, 30*(16), 1799-1807.

Bruce, J., & Quinlan, J. (2011). Chronic Post Surgical Pain. *Reviews in Pain, 5*(3), 23-29. doi:10.1177/204946371100500306

Budtz-Lilly, A., Fink, P., Ørnbøl, E., Vestergaard, M., Moth, G., Christensen, K. S., & Rosendal, M. (2015). A new questionnaire to identify bodily distress in primary care: The 'BDS checklist'. *Journal of Psychosomatic Research, 78*(6), 536-545. doi:10.1016/j.jpsychores.2015.03.006

Campbell, M. K., Fayers, P. M., & Grimshaw, J. M. (2005). Determinants of the intracluster correlation coefficient in cluster randomized trials: The case of implementation research. *Clinical Trials: Journal of the Society for Clinical Trials, 2*(2), 99-107. doi:10.1191/1740774505cn071oa

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods. *Advances in Methods and Practices in Psychological Science, 2*(2), 115-144. doi:10.1177/2515245919847196

Claassen-van Dessel, N., van der Wouden, J. C., Hoekstra, T., Dekker, J., & van der Horst, H. E. (2018). The 2-year course of Medically Unexplained Physical Symptoms (MUPS) in terms of symptom severity and functional status: results of the PROSPECTS cohort study. *J Psychosom Res, 104*, 76-87. doi:10.1016/j.jpsychores.2017.11.012

Creed, F., Ratcliffe, J., Fernandes, L., Palmer, S., Rigby, C., Tomenson, B., . . . Thompson, D. G. (2005). Outcome in severe irritable bowel syndrome with and without accompanying depressive, panic and neurasthenic disorders. *British Journal of Psychiatry, 186*(6), 507-515. doi:10.1192/bjp.186.6.507

del Pozo-Cruz, B., Parraca, J. A., del Pozo-Cruz, J., Adsuar, J. C., Hill, J., & Gusi, N. (2012). An occupational, internet-based intervention to prevent chronicity in subacute lower back pain: a randomised controlled trial. *Journal of Rehabilitation Medicine, 44*(7), 581-587. doi:10.2340/16501977-0988

DerSimonian, R., & Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials, 28*(2), 105-114. doi:10.1016/j.cct.2006.04.004

De Waal, M. W. M., Arnold, I. A., Eekhof, J. A. H., & Van Hemert, A. M. (2004). Somatoform disorders in general practice. Prevalence, functional impairment and comorbidity with anxiety and depression. *British Journal of Psychiatry, 184*(6), 470-476. doi:10.1192/bjp.184.6.470

Gatchel, R. J., Stowell, A. W., Wildenstein, L., Riggs, R., & Ellis, E., 3rd. (2006). Efficacy of an early intervention for patients with acute temporomandibular disorder-related pain: a one-year outcome study. *Journal of the American Dental Association, 137*(3), 339-347. doi:10.14219/jada.archive.2006.0183

Hausteiner-Wiehle, C., Schäfert, R., Häuser, W., Herrmann, M., Ronel, J., Sattel, H., & Henningsen, P. (2012). S3- Leitlinie Umgang mit Patienten mit nicht-spezifischen, funktionellen und somatoformen Körperbeschwerden. [Guideline for the treatment of patients with non-specific, functional and somatoform physical complaints]. AWMF-Reg.-Nr. 051–001 2012; www.awmf.org/leitlinien/detail/ll/051-001.html. Retrieved 12.03.2015

Hedges, L. V. (1981). Distribution Theory for Glasss Estimator of Effect Size and Related Estimators.

Journal of Educational Statistics, 6(2), 107. doi:10.2307/1164588

Henningsen, P., Gündel, H., Kop, W. J., Löwe, B., Martin, A., Rief, W., . . . Euronet-Soma Group. (2018). Persistent physical symptoms as perceptual dysregulation: A neuropsychobehavioral model and its clinical implications. *Psychosomatic Medicine, 80*(5), 422-431.

Henningsen, P., Zimmermann, T., & Sattel, H. (2003). Medically Unexplained Physical Symptoms, Anxiety, and Depression. *Psychosomatic Medicine, 65*(4), 528-533. doi:10.1097/01.psy.0000075977.90337.e7

Henningsen, P., Zipfel, S., Sattel, H., & Creed, F. (2018). Management of Functional Somatic Syndromes and Bodily Distress. *Psychotherapy and Psychosomatics, 87*(1), 12-31. doi:10.1159/000484413

Herzog, A., Shedden-Mora, M. C., Jordan, P., & Löwe, B. (2018). Duration of untreated illness in patients with somatoform disorders. *Journal of Psychosomatic Research, 107*, 1-6. doi:10.1016/j.jpsychores.2018.01.011

Higgins, J. P. T., Deeks, J., & Altman, D. G. (2011). Special topics in statistics. In J. P. T. Higgins & S. Green (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.* The Cochrane Collaboration. Retrieved from http://handbook-5-1.cochrane.org/

Higgins, J. P. T., & Green, S. (Eds.) (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.* The Cochrane Collaboration. Retrieved from http://handbook-5-1.cochrane.org/

Higgins, J. P. T., Savović, J., Page, M. J., Sterne, J. A. C., on behalf of the ROB2 Development Group (2019). Revised Cochrane risk-of-bias tool for randomized trials (RoB 2). Retrieved from https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool/current-version-of-rob-2

Inthout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology, 14*(1). doi:10.1186/1471-2288-14-25

Inthout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open, 6*(7). doi:10.1136/bmjopen-2015-010247

Karimi, M., & Brazier, J. (2016). Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? *PharmacoEconomics, 34*(7), 645-649. doi:10.1007/s40273-016-0389-9

Kisely, S., Scott, A., Denney, J., & Simon, G. (2006). Duration of untreated symptoms in common mental disorders: Association with outcomes. *British Journal of Psychiatry, 189*(1), 79-80. doi:10.1192/bjp.bp.105.019869

Kleinstäuber, M., Witthöft, M., & Hiller, W. (2011). Efficacy of short-term psychotherapy for multiple medically unexplained physical symptoms: A meta-analysis. *Clinical Psychology Review, 31*, 146-160. doi:http://dx.doi.org/10.1016/j.cpr.2010.09.001

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22*(17), 2693-2710. doi:10.1002/sim.1482

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2002). The PHQ-15: Validity of a New Measure for Evaluating the Severity of Somatic Symptoms. *Psychosomatic Medicine, 64*(2), 258-266. doi:10.1097/00006842-200203000-00008

Koelen, J. A., Houtveen, J. H., Abbass, A., Luyten, P., Eurelings-Bontekoe, E. H., Van Broeckhuysen-Kloth, S. A., . . . Geenen, R. (2014). Effectiveness of psychotherapy for severe somatoform disorder: meta-analysis. *British Journal of Psychiatry, 204*(1), 12-19. doi:10.1192/bjp.bp.112.121830

Konnopka, A., Schaefert, R., Heinrich, S., Kaufmann, C., Luppa, M., Herzog, W., & König, H. H. (2012). Economics of medically unexplained symptoms: A systematic review of the literature. *Psychotherapy and Psychosomatics, 81*, 265-275. doi:10.1159/000337349

Kroenke, K. (2007). Efficacy of treatment for somatoform disorders: A review of randomized controlled trials. *Psychosomatic Medicine, 69*, 881-888.

Langan, D., Higgins, J. P., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., . . . Simmonds, M. (2018). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods, 10*(1), 83-98. doi:10.1002/jrsm.1316

Lefebvre, C., Manheimer, E., & Glanville, J. (2011). Searching for studies. In Higgins, J. P. T., & Green, S. (Eds.) (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.* The Cochrane Collaboration. Retrieved from http://handbook-5-1.cochrane.org/

Löwe, B., Lau, K., Daubmann, A., Härter, M., Wegscheider, K., König, H.-H., & Shedden-Mora, M. C. (2017). Effectiveness of a stepped, collaborative, and coordinated health care network for somatoform disorders (Sofu-Net): A controlled cluster cohort study. *Psychosomatic Medicine, 79*(9), 1016-1024.

Löwe, B., Lohse, A., Andresen, V., Vettorazzi, E., Rose, M., & Broicher, W. (2016). The Development of Irritable Bowel Syndrome: A Prospective Community-Based Cohort Study. *The American Journal of Gastroenterology, 111*(9), 1320-1329. doi:10.1038/ajg.2016.255

Löwe, B., Spitzer, R. L., Williams, J. B., Mussell, M., Schellberg, D., & Kroenke, K. (2008). Depression, anxiety and somatization in primary care: Syndrome overlap and functional impairment. *General Hospital Psychiatry, 30*(3), 191-199. doi:10.1016/j.genhosppsych.2008.01.001

Moons, P. (2004). Why Call it Health-Related Quality of Life When You Mean Perceived Health Status? *European Journal of Cardiovascular Nursing, 3*(4), 275-277. doi:10.1016/j.ejcnurse.2004.09.004

Nimnuan, C., Hotopf, M., & Wessely, S. (2001). Medically unexplained symptoms. An epidemiological survey in seven specialities. *Journal of Psychosomatic Research, 51*(1), 361-367. doi:10.1016/s0022-3999(01)00223-9

Okuda, A., Suzuki, T., Kishi, T., Yamanouchi, Y., Umeda, K., Haitoh, H., . . . Iwata, N. (2010). Duration of untreated illness and antidepressant fluvoxamine response in major depressive disorder. *Psychiatry and Clinical Neurosciences, 64*(3), 268-273. doi:10.1111/j.1440-1819.2010.02091.x

Oliveira, A., Gevirtz, R., & Hubbard, D. (2006). A psycho-educational video used in the emergency department provides effective treatment for whiplash injuries. *Spine, 31*(15), 1652-1657.

Petersen, M. W., Schroder, A., Jorgensen, T., Ornbol, E., Dantoft, T. M., Eliasen, M., . . . Fink, P. (2019). Prevalence of functional somatic syndromes and bodily distress syndrome in the Danish population: the DanFunD study. *Scand J Public Health*, 1403494819868592. doi:10.1177/1403494819868592

R Core Team (2019). R: A language for statistical computing [Computer software]. Retrieved from https://www.R-project.org/

Rief, W., Burton, C., Frostholm, L., Henningsen, P., Kleinstäuber, M., Kop, W. J., . . . Euronet-Soma Group. (2017). Core Outcome Domains for Clinical Trials on Somatic Symptom Disorder, Bodily Distress Disorder, and Functional Somatic Syndromes: European Network on Somatic Symptom Disorders Recommendations. *Psychosomatic Medicine, 79*(9), 1008-1015. doi:10.1097/PSY.0000000000000502

Roenneberg, C., Hausteiner-Wiehle, C., Schäfert, R., Sattel, H., & Henningsen, P. (2018). Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF): S3 Leitlinie "Funktionelle Körperbeschwerden". Retrieved from https://www.awmf.org/uploads/tx_szleitlinien/051-001l_S3_Funktionelle_Koerperbeschwerden_2018-11.pdf

Rosendal, M., Olesen, F., Fink, P., Toft, T., Sokolowski, I., & Bro, F. (2007). A randomized controlled trial of brief training in the assessment and treatment of somatization in primary care: effects on patient outcome. *General Hospital Psychiatry, 29*(4), 364-373. doi:10.1016/j.genhosppsych.2007.03.005

Schaefert, R., Laux, G., Kaufmann, C., Schellberg, D., Bolter, R., Szecsenyi, J., . . . Kuehlein, T. (2010). Diagnosing somatisation disorder (P75) in routine general practice using the International Classification of Primary Care. *Journal of Psychosomatic Research, 69*, 267-277. doi:10.1016/j.jpsychores.2010.05.003

Schünemann, H., Brożek, J., Guyatt, G., & Oxman, A. (Eds.) (2013). *GRADE Handbook.* Retrieved from https://gdt.gradepro.org/app/handbook/handbook.html

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ, 349*(Jan02 1). doi:10.1136/bmj.g7647

Shedden-Mora, M. C., Groß, B., Lau, K., Gumz, A., Wegscheider, K., & Löwe, B. (2016). Collaborative stepped care for somatoform disorders: A pre–post-intervention study in primary care. *Journal of Psychosomatic Research, 80*, 23-30. doi:10.1016/j.jpsychores.2015.11.004

Smith, K. W., Avis, N. E., & Assmann, S. F. (1999). Distinguishing between quality of life and health status in quality of life research: A meta-analysis. *Quality of Life Research, 8,* 447-459.

Spiegel, B. M., Gralnek, I. M., Mayer, E. A., Bolus, R., Chang, L., Dulai, G. S., & Naliboff, B. (2003). Clinical determinants of health-related quality of life in irritable bowel syndrome. *Gastroenterology, 124*(4). doi:10.1016/s0016-5085(03)82014-0

Stanley, T. D., & Doucouliagos, H. (2013). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*(1), 60-78. doi:10.1002/jrsm.1095

Testa, M. A., & Simonson, D. C. (1996). Assessment of quality-of-life outcomes. *New England Journal of Medicine, 334,* 835-840. doi:10.1056/NEJM199603283341306

Thabane, M., & Marshall, J. K. (2009). Post-infectious irritable bowel syndrome. *World Journal of Gastroenterology, 15*(29), 3591. doi:10.3748/wjg.15.3591

Toft, T., Rosendal, M., Ornbol, E., Olesen, F., Frostholm, L., & Fink, P. (2010). Training general practitioners in the treatment of functional somatic symptoms: Effects on patient health in a cluster-randomised controlled trial (the Functional Illness in Primary Care study). *Psychotherapy and Psychosomatics, 79*(4), 227-237. doi:10.1159/000313691

Van Dessel, N., Den Boeft, M., Van Der Wouden, J. C., Kleinstäuber, M., Leone, S. S., Terluin, B., . . . Van Marwijk, H. (2014). Non-pharmacological interventions for somatoform disorders and medically unexplained physical symptoms (MUPS) in adults. *Cochrane Database of Systematic Reviews*. doi:10.1002/14651858.cd011142.pub2

van Gils, A., Schoevers, R. A., Bonvanie, I. J., Gelauff, J. M., Roest, A. M., & Rosmalen, J. G. (2016). Self-Help for Medically Unexplained Symptoms: A Systematic Review and Meta-Analysis. *Psychosomatic Medicine, 78*(6), 728-739. doi:10.1097/PSY.0000000000000325

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika, 60*(3), 419-435. doi:10.1007/bf02294384

von Elm, E., Poglia, G., Walder, B., & Tramèr, M. R. (2004). Different Patterns of Duplicate Publication. *JAMA, 291*(8), 974. doi:10.1001/jama.291.8.974

Ware, J. E. (2000). SF-36 Health Survey Update. *Spine, 25*(24), 3130-3139. doi:10.1097/00007632-200012150-00008

Wessely, S., Nimnuan, C., & Sharpe, M. (1999). Functional somatic syndromes: one or many? *The Lancet, 354*(9182), 936-939.

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jonsson, B., . . . Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology, 21*(9), 655-679. doi:10.1016/j.euroneuro.2011.07.018

Appendix to study protocol: List of functional disorders & search strategy

Clinical conditions

Specialty	Condition	Search terms
General terms		*Functional disorder* *Functional disorders* *Functional symptom* *Functional symptoms* *Functional syndrome* *Functional syndromes* *Functional somatic* *Functional illness* *Functional illnesses* *Idiopathic* *Non-specific* *Non-organic* *Psychogenic* *Dissociative* *Medically unexplained* *Organically unexplained* *unexplained* *Psychosomatic* *Somatoform* *Persistent physical symptoms* *Persistent somatic symptoms* *Mimic* *Mimics* *Mimick* *Mimicks*
	Multisomatoform disorder	*Multisomatoform*
	Somatization disorder	*Somatization* *Somatisation* *Briquet syndrome*
	Pain disorder	*Pain disorder* *Pain disorders*
	Conversion disorder	*Conversion* *Hysteria* *Hysteric*
	Somatic symptom disorder	*Somatic symptom disorder* *Somatic symptom disorders* *Somatic symptom distress*
	Bodily distress disorder	*Bodily distress* *Body distress*
	Bodily stress syndrome	*Bodily stress* *Body stress*
	Neurasthenia	*Neurasthenia*
	Culture-bound syndrome	*Culture-bound* *Culture-specific*
Allergology		
	Food intolerance	Food intolerance Food sensitivity Food sensitivities

		Food hypersensitivity Food hypersensitivities Food allergy Food allergies Pseudo-allergy Pseudo-allergies Functional food intolerance Functional food sensitivity Functional food sensitivities Functional food hypersensitivity Functional food hypersensitivities Functional food allergy Functional food allergies
	Multiple chemical sensitivity	chemical sensitivity chemical sensitivities Idiopathic environmental
	Sick building syndrome	Sick building Sick house
	Persian gulf syndrome	Persian gulf syndrome Gulf war syndrome
	Amalgam hypersensitivity	Amalgam hypersensitivity Dental amalgam Dental Amalgam toxicity Functional amalgam hypersensitivity Functional amalgam toxicity
	implant intolerance	Implant intolerance
	Prosthesis intolerance	Prosthesis intolerance
	Aerotoxic syndrome	Aerotoxic Sick aeroplane
Anesthesiology		
	Idiopathic pain	Idiopathic pain Panalgesia Psychogenic pain Functional pain Unspecific pain
	Chronic postoperative pain	Chronic postoperative
Cardiology		
	Atypical chest pain	Atypical chest pain Nonspecific chest pain Non-specific chest pain Noncardiac chest pain Non-cardiac chest pain Functional chest pain
	Palpitations with normal investigations	Psychogenic palpitation Psychogenic palpitations Functional palpitation Functional palpitations
	Syndrome X	Syndrome X Syndrome Xs Microvascular angina
Dermatology		
	Psychogenic skin disease	Psychogenic skin disease

		Psychogenic skin diseases Psychogenic pruritus Functional skin
Endocrinology		
	Hypoglycaemia	Psychogenic hypoglycaemia Psychogenic hypoglycemia Idiopathic postprandial syndrome Functional hypoglycaemia Functional hypoglycemia
Gastroenterology		*Functional gastrointestinal*
	Functional bowel disorders	Functional bowel
	Irritable bowel syndrome	Irritable bowel Irritable colon IRS Mucous colitis Mucous colitides
	Nonulcer dyspepsia	Nonulcer dyspepsia Functional dyspepsia
	Functional Abdominal pain	Functional abdominal Psychogenic abdominal
	Functional colonical disease	Functional colonical
	Functional disorders of swallowing	Functional swallowing Psychogenic dysphagia Globus sensation Globus sensations
Gynecology		
	Premenstrual syndrome	Premenstrual syndrome Premenstrual syndromes Premenstrual dysphoria Premenstrual dysphoric PMDD Late luteal phase dysphoria Late luteal phase dysphoric Premenstrual tension Premenstrual tensions
Infectiology		
	Chronic lyme disease	Chronic lyme
	Candida hypersensitivity	Candida hypersensitivity Candidas hypersensitivity Candida syndrome
	Chronic rhinopharyngitis	Chronic rhinopharyngitis
Neurology		*Functional neurologic* *Functional neurological* *General functional neurologic* *General functional neurological* *Mixed functional neurologic* *Mixed functional neurological*
	Functional seizures	Functional seizure Functional seizures Non-epileptic seizure Non-epileptic seizures PNES

		Pseudoseizure Pseudoseizures Pseudo-seizure Pseudo-seizures Hysterical seizure Hysterical seizures Non-epileptic attack Non-epileptic attacks Dissociative seizure Dissociative seizures Dissociative attack Dissociative attacks
	Functional voice disorder	Functional voice Functional dysphonia Functional aphonia Muscle tension voice disorder Muscle tension voice disorders Psychogenic voice
	Functional motor disorder Functional movement disorder Functional sensorimotor disorder	Functional motor Functional movement Functional sensorimotor Functional weakness Functional weaknesses Functional leg Functional limb Functional arm Functional paralysis Functional tremor Functional dystonia Posttraumatic painful torticollis Functional jerk Functional jerks Functional tic Functional tics Functional myoclonus Functional paroxysmal Functional gait Movement disorder mimic Movement disorder mimics Neurologic mimic Neurologic mimics Musculoskeletal mimic Musculoskeletal mimics Biomechanical mimic Biomechanical mimics Isolated disequilibrium Functional balance Functional parkinsonism
	Functional eye movement disorder	Functional eye Functional convergence spasm Functional convergence spasms Functional convergence paralysis

		Functional gaze limitation Functional gaze limitations Functional eye oscillation Functional eye oscillations Functional nystagmus Functional opsoclonus Functional tonic eye deviation Functional tonic eye deviations Functional oculogyric crisis Functional diplopia Functional tonic gaze deviation Functional tonic gaze deviations
	Functional facial movement disorder Functional tongue movement disorder	Functional facial Functional tongue Psychogenic blespharospasm Functional blespharospasm Functional oromandibular dystonia Functional facial dystonia
	Functional sensory symptoms	Functional sensory Functional hypoesthesia Functional Hyperesthesia Functional Hemihyperesthesia Functional Paresthesia
	Functional visual symptoms	Functional visual Functional visual loss
	Functional auditory disorders	Functional auditory Functional hearing loss Auditory processing disorder Auditory processing disorders Tinnitus Low-frequency noise complaint Low-frequency noise complaints Infrasound hypersensitivity Sound tolerance Loudness perception Hyperacusis Misophonia Acoustic shock Acoustic shocks
	Functional speech disorder	Functional speech Functional stuttering Functional dysfluency Functional articulation Prosodic abnormality Prosodic abnormalities Foreign accent syndrome Foreign accent syndromes Abnormal resonance Hypernasality
	Functional memory disorder	Functional memory

	Functional cognitive disorder	Functional cognitive Functional amnesia
	Functional dizziness	Functional dizziness Dizziness Phobic postural vertigo Chronic subjective dizziness CSD Persistent postural-perceptual dizziness PPPD Subjective dizziness Chronic dizziness Persistent dizziness
	Functional stroke	Functional stroke Stroke mimic Stroke mimics
	Tension headache	Tension headache Tension headaches Tension-type headache Tension-type headaches Tension type headache Tension type headaches Tension-vascular headache Tension-vascular headaches Tension vascular headache Tension vascular headaches TTH Stress headache Stress headaches Functional headache Functional headaches
	Atypical face pain	Atypical face pain Facial pain Myofacial pain Functional face pain Functional facial pain
	Electromagnetic hypersensitivity	Electromagnetic hypersensitivity Electro-hypersensitivity Electrosensitivity Electro-sensitivity Electricity hypersensitivity IEI-EMF Environmental illness Environmental illnesses
	Central sensitivity syndrome	Central sensitivity
	Post-concussion syndrome	Post-concussion Post concussion Post-concussive Post concussive PCS Post-traumatic complaints
Oral medicine / Otorhinolaryngology		

	Temporomandibular joint disorder	Temporomandibular joint Temporo-mandibular joint Temporomandibular disorder Temporomandibular disorders Temporo-mandibular disorder Temporo-mandibular disorders Temporomandibular dysfunction Temporomandibular dysfunctions Temporo-mandibular dysfunction Temporo-mandibular dysfunctions TMJ TMJD Craniomandibular disorder Craniomandibular disorders Cranio-mandibular disorder Cranio-mandibular disorders
	Atypical odontalgia	Atypical odontalgia Atypical odontalgias Functional odontalgia Functional odontalgias
	Psychogenic gagging	Psychogenic gagging Functional gagging
	Burning mouth	Burning mouth Glossalgia Glossalgias Glossodynia Glossodynias Glossopyrosis Glossopyroses
	Bruxism	Bruxism
	Globus syndrome	Globus syndrome Globus syndromes Globus hystericus Globus pharynges
Orthopedics		
	Repetitive strain injury	Repetitive strain Repetition strain Overuse injury Overuse injuries Overuse syndrome Overuse syndromes Repetitive stress Repetitive motion Cumulative trauma disorder Cumulative trauma disorders
	Chronic whiplash syndrome	Chronic whiplash Whiplash associated Whiplash-associated
	Neck pain	Neck pain Chronic neck pain Functional neck pain
Respiratory		

Medicine		
	Hyperventilation syndrome	Hyperventilation syndrome Hyperventilation syndromes
Rheumatology		*Functional rheumatologic* *Functional rheumatological*
	Fibromyalgia	Fibromyalgia FMS Chronic widespread pain Widespread musculoskeletal pain Myofascial pain
	Chronic low back pain	Nonspecific back pain Non-specific back pain Lower back pain Low back pain Functional back pain
	Chronic pain Persistent pain Chronic intractable benign pain syndrome	Chronic pain Persistent pain Chronic intractable benign pain CIBPS
	Chronic fatigue syndrome Myalgic encephalomyelitis Post-viral fatigue syndrome	Chronic fatigue Myalgic encephalomyelitis Post-viral fatigue postviral fatigue post viral fatigue myalgic encephalopathy chronic epstein barr virus chronic Epstein-barr virus chronic mononucleosis chronic infectious mononucleosis like chronic fatigue and immune effort syndrome effort syndromes low natural killer cell syndrome low natural killer cell syndromes neuromyasthenia postviral syndrome postviral syndromes post-viral syndrome post-viral syndromes post viral syndrome post viral syndromes post infectious fatigue postinfectious fatigue post-infectious fatigue Fatigue syndrome Fatigue syndromes Psychogenic fatigue systemic exertion intolerance CFS ME ME/CFS
Urology		

	Functional urologic disorders	Functional urologic Functional urinary Functional micturition Micturition dysfunction Micturition dysfunctions
	Fowler's syndrome	Fowler's syndrome Psychogenic urinary retention Functional urinary retention
	Paruresis	Paruresis Shy-bladder Shy bladder Bashful bladder
	Dysfunctional voiding	Dysfunctional voiding Hinman-Allen Hinman Nonneurogenic neurogenic bladder Non-neurogenic neurogenic bladder
	Idiopathic overactive bladder	Idiopathic overactive bladder Irritable bladder
	Interstitial cystitis	Interstitial cystitis Interstitial cystitides Bladder pain Painful bladder
	Urethral syndrome	Urethral syndrome Urethral syndromes
	Chronic pelvic pain syndrome	Pelvic pain CPPS Unspecific pelvic pain Unexplained pelvic pain
	Pelvic arthropathy	Pelvic arthropathy

Note: Functional coma, incl. functional stupor & non-epileptic pseudo-status epilepticus, as well as pseudocyesis (false pregnancy) not included in the search terms, since early psychological interventions make no sense conceptually. Factitious disorder excluded.

Cave: Food intolerance / sensitivity not always functional. Needs to be considered when selecting studies.

Early psychological interventions

Function	**Search terms**
Focusing search on early interventions	Early New onset Recent onset Sub-acute Acute Sub-threshold Sub-clinical Non-chronic Psychological first aid
Focusing search on preventive interventions	Prevent Preventary Preventive Preventative Preventing Prevention
Focusing search on (psychological) interventions	Intervention Interventions Therapy Therapeutic Treatment Treatments Management Psychotherapy Psychotherapeutic CBT Psychoeducation Psychoeducational Psycho-education Psycho-educational Education Educational Self-help Self help Information Rehabilitation Bibliotherapy Bibliotherapeutic

Search will be conducted for each functional condition, separately. Search terms will be combined using the Boolean operator "OR". Search terms for each condition will be combined with the following search phrase intended to narrow the search on early psychological interventions using the Boolean operator "AND":

(Early intervention OR Early interventions OR Early therapy OR Early therapeutic OR Early treatment OR Early treatments OR Early management OR Early psychotherapy OR Early psychotherapeutic OR Early CBT OR Early psychoeducation OR Early psychoeducational OR Early psycho-education OR Early psycho-educational OR Early education OR Early educational OR Early self-help OR Early self help OR

Early information OR Early rehabilitation OR Early bibliotherapy OR Early bibliotherapeutic OR ((new onset OR recent onset OR sub-acute OR acute OR sub-threshold OR sub-clinical OR non-chronic) AND (intervention OR interventions OR therapy OR treatment OR treatments OR management OR psychotherapy OR CBT OR psychoeducation* OR psycho-education* OR education* OR self-help OR self help OR information OR rehabilitation OR bibliotherap*)) OR prevent OR preventary OR preventive OR preventative OR preventing OR prevention OR psychological first aid)

Search will be limited to titles and abstracts of records. Additionally, we will employ filters for detecting randomized controlled trials and studies published from 1994 until 1st September 2019, only. For all compound search terms, phrase searching will be conducted.

References for search terms:

Barnett, C., Armes, J., & Smith, C. (2019). Speech, language and swallowing impairments in functional neurological disorder: a scoping review. *International Journal of Language & Communication Disorders, 54*(3), 309-320. doi:10.1111/1460-6984.12448

Espay, A. J., Aybek, S., Carson, A., Edwards, M. J., Goldstein, L. H., Hallett, M., . . . Morgante, F. (2018). Current Concepts in Diagnosis and Treatment of Functional Neurological Disorders. *JAMA Neurology, 75*(9), 1132-1141. doi:10.1001/jamaneurol.2018.1264

Fink, P. (Producer). (2017). Syndromes of bodily distress or functional somatic syndromes - where are we heading? Lecture on occasion of receiving the Alison Creed Award 2017. [Presentation] Retrieved from http://eapm2017.com/images/site/abstracts/PLENARY_Prof_FINK.pdf

Hallett, M., Stone, J., & Carson, A. (Eds.) (2016). Functional neurologic disorders. In M. J. Aminoff, F., Boller, & D. F. Swaab (Series Eds.), *Handbook of clinical neurology (Vol. 139, 3rd series).* Amsterdam: Elsevier.

Henningsen, P., Zipfel, S., Sattel, H., & Creed, F. (2018). Management of Functional Somatic Syndromes and Bodily Distress. *Psychotherapy and Psychosomatics, 87*(1), 12-31. doi:10.1159/000484413

Kleinstauber, M., Witthoft, M., & Hiller, W. (2011). Efficacy of short-term psychotherapy for multiple medically unexplained physical symptoms: a meta-analysis. *Clinical Psychology Review, 31*(1), 146-160. doi:10.1016/j.cpr.2010.09.001

Kroenke, K., & Swindle, R. (2000). Cognitive-Behavioral Therapy for Somatization and Symptom Syndromes: A Critical Review of controlled clinical trials. *Psychotherapy and Psychosomatics, 69*, 205-215.

Ludwig, L., Pasman, J. A., Nicholson, T., Aybek, S., David, A. S., Tuck, S., . . . Stone, J. (2018). Stressful life events and maltreatment in conversion (functional neurological) disorder: systematic review and meta-analysis of case-control studies. *The Lancet Psychiatry, 5*(4), 307-320. doi:10.1016/s2215-0366(18)30051-8

Roenneberg, C., Hausteiner-Wiehle, C., Schäfert, R., Sattel, H., & Henningsen, P. (2018). Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF): S3 Leitlinie "Funktionelle Körperbeschwerden": Leitlinienreport. Retrieved from https://www.awmf.org/uploads/tx_szleitlinien/051-001l_S3_Funktionelle_Koerperbeschwerden_2018-11.pdf

Teodoro, T., Edwards, M. J., & Isaacs, J. D. (2018). A unifying theory for cognitive abnormalities in functional neurological disorders, fibromyalgia and chronic fatigue syndrome: systematic review. *Journal of Neurology, Neurosurgery & Psychiatry, 89*(12), 1308-1319. doi:10.1136/jnnp-2017-317823

van Gils, A., Schoevers, R. A., Bonvanie, I. J., Gelauff, J. M., Roest, A. M., & Rosmalen, J. G. (2016). Self-Help for Medically Unexplained Symptoms: A Systematic Review and Meta-Analysis. *Psychosomatic Medicine, 78*(6), 728-739. doi:10.1097/PSY.0000000000000325

Online Supplement B

Amendments to the protocol

Table B1

Description of amendments to the protocol

Date	Section	Amendment	Rationale
23th August 2019	Study records	The study selection process will be aided by a student assistant instead of MSM. The student assistant will be introduced to the subject matter and the procedure by LB.	We decided to conduct the study selection process together with a student assistant, since MSM had no available resources for the study selection process and we considered the student assistant to be competent and knowledgeable enough for this task, especially after a personal introduction to the subject.
10th September 2019	Search strategy	PsycINFO search: The filter "Therapy (best balance of sensitivity and specificity)" was used and search was limited to appropriate publication years without specification of months. Mapping of terms to subject headings was deactivated.	PsycINFO does not offer a "Randomized Controlled Trial" filter and does not allow to specify months when filtering for publication date. Thus, we employed the "Therapy" filter and publication year was limited to 1994 – 2019. Mapping of subject headings was deactivated, because our search strategy extensively covered search terms of interest already.

Table B1 (*Continued*)

Date	Section	Amendment	Rationale
10th September 2019	Search strategy	„IBS“ added to search terms for irritable bowel syndrome.	We added the search term „IBS“ since this is a commonly used abbreviation for irritable bowel syndrome.
11th September 2019	Search strategy	Web of Science search: The search was limited to appropriate publication years without specification of months. The filter „Title“ was used.	Web of Science does not allow to specify months when filtering for publication date. Thus, publication year was limited to 1994 – 2019. The filter „Title“ was used because using the filter „Title/Abstract“ led to an inflation of search results (at least 30 000 records) mirroring low specificity and overconsuming our resources.
12th September 2019	Study records	Deduplication was conducted according to the algorithm proposed by Bramer, Giustini, de Jonge, Holland, and Bekhuis (2016).	We detected the paper of Bramer et al. (2016) and conceived the proposed procedure for deduplication to be more practical than using the built-in deduplication function in EndNote.
01st October 2019	Study records	Title-abstract screening was conducted by LB, only.	Title-abstract screening was conducted by LB, only, due to illness of the student assistant.

Table B1 (*Continued*)

Date	Section	Amendment	Rationale
08th October 2019	Study records	For all abstracts in the full-text screening, we looked for a full text publication of the respective data. When the data were published in a report, we included the respective report in the full-text screening. Otherwise, the abstract was excluded. Editorials, letters and other grey literature were also excluded.	During full-text screening, we were confronted with abstracts and other grey literature. Handling grey literature was not specified before. The decision for this procedure was made before full-text screening was finished.
22nd October 2019	Eligibility criteria	When psychological interventions are combined with other interventions, the volume of the psychological intervention must be at least as large as the volume of the other interventions.	During full-text screening, we were confronted with multidisciplinary intervention studies. Handling of such studies was not specified before.
22th October 2019	Study records	After having finished the selection process, we will look for errata, corrections etc. of included studies.	We entered this step to ensure integrity of included data. The decision to implement this procedure was made before full-text screening was finished.

Table B1 (*Continued*)

Date	Section	Amendment	Rationale
09th November 2019	Data items	Length of follow-up measurement was calculated in reference to the end of treatment. If length of follow-up differed between subjects within a study, we coded the shortest length of follow-up.	Treatments in the included studies were different in length. As we specified moderator analyses of length of follow-up in our protocol, we coded length of follow-up in reference to the end of treatment (instead of, for example, start of the trial) in order to account for these differences in lengths of treatments. We coded the shortest length of follow-up in cases where length of follow-up varied between subjects to obtain a conservative estimate. These decisions were made prior to data analysis.
10th December 2019	Data collection process	Difficult coding decisions will be discussed with MSM.	Originally, data collection was planned to be conducted by LB, only. As we noticed many difficult coding decisions, we decided to discuss these decisions together. The decision to implement this procedure was made after data collection, but before data analysis.

Table B1 (*Continued*)

Date	Section	Amendment	Rationale
25th March 2020	Risk of bias in individual studies	Risk of bias in cluster-randomized trials will be assessed using the Revised Cochrane risk of bias tool for randomized trials (RoB 2.0; Eldridge et al., 2016).	We decided to employ the Revised Cochrane risk of bias tool for randomized trials ((RoB 2.0); Eldridge et al., 2016) instead of the newer version of this tool (Revised Cochrane risk-of-bias tool for randomized trials (RoB 2); Sterne et al., 2019), since the latter is not suitable for cluster-randomized trials. The decision for employing the RoB 2.0 was made after main data analyses, but before moderator and sensitivity analyses.
04th April 2020	Additional analyses	The relationship between intervention intensity and ordinal variables (type of participant population, type of control group) as well as among these ordinal variables will be examined using Cramer's V. The relationship between ordinal and metric variables (duration of symptoms, length of follow-up) will be examined using Spearman's rho (ρ). The relationship between intervention intensity and metric variables will be assessed using biserial correlation, while the relationships between metric variables will be quantified via Pearson's r.	In our protocol, we did not specify by which statistics the interrelations between our moderator variables shall be determined. Therefore, we specified these statistics prior to running moderator analyses.

Table B1 (*Continued*)

Date	Section	Amendment	Rationale
04th April 2020	Sensitivity analyses	We will explore the effect of including cluster-randomized trials on our results by rerunning the analyses without cluster-randomized trials.	Originally, we planned to conduct this sensitivity analyses by including trial design as moderator variable. However, we think that simply excluding cluster-randomized trials is sufficient for our purposes, especially when considering the low amount of cluster-randomized trials in this review. The decision for this amendment was made before starting with moderator and sensitivity analyses.
15th April 2020	Years considered, Information sources, Search strategy	We will update our search by repeating our electronic database search in order to detect records published between 01st September 2019 and 30th April 2020.	After finishing the statistical analyses, we scrutinized publication requirements of different journals and noticed that some journals demand more current data than ours. Therefore, we decided to update our search and our analyses. This decision was unrelated to the interim results of our statistical analyses.

Table B1 (*Continued*)

Date	Section	Amendment	Rationale
29th April 2020	Confidence in cumulative estimate	We will not conduct a structured GRADE assessment of our findings.	We revised our decision to conduct a GRADE assessment of the evidence since we noticed that a GRADE assessment demands clear and specific review questions which is contrary to the rationale of this review. Furthermore, it would have been indicated to specify a prioritization of outcomes and comparisons in advance which we did not conduct.

Online Supplement C

Outcome characteristics of included studies

Table C1

Outcome data of studies measuring somatic symptom severity

Study[a]	Measure	high value desirable?		baseline		post-treatment			follow-up			
				M (*SD*)	*n*	*M* (*SD*)	*n*	*g* [95%-CI]	length of follow-up (months)[b]	*M* (*SD*)	*n*	*g* [95%-CI]
Bérubé et al., 2019	BPI average pain intensity upon movement in last 7 days	no	Intervention:	6.3 (2.3)	28	2.3 (2.2)	25	0.046 [-0.51, 0.61]	3	2.5 (2)	25	-0.18 [-0.75, 0.36]
			Control:	6.2 (2.4)	28	2.4 (2.1)	24			2.1 (2.4)	23	
Birch et al., 2020[c]	VAS pain during activity	no	Intervention:	48 (18)	31	22 (18.9)	24	-0.38 [-0.95, 0.2]	9	12 (15.4)	24	-0.2 [-0.75, 0.36]
			Control:	49 (21)	29	15 (17.8)	24			9 (14.9)	26	
Bjørnnes et al., 2017	BPI-SF	no	Intervention:	-	-	3.9 (2.3)	174	-0.32 [-0.53, -0.11]	12	1.2 (2.1)	174	0 [-0.21, 0.21]
			Control:	-	-	3.2 (2.1)	175			1.2 (2)	175	
Cai et al. 2018	11-point NRS (knee pain)	no	Intervention:	6.64 (1.03)	50	-	-	-	6	5.63 (0.73)	50	0.8 [0.39, 1.2]
			Control:	6.71 (1.17)	50	-	-			6.27 (0.86)	50	
Dahl & Nilsson, 2001	MPI pain severity subscale	no	Intervention:	2.11 (1.22)	NA	-	-	-	12	2.63 (0.88)	NA	NA
			Control:	2.07 (1.45)	NA	-	-			2.43 (0.86)	NA	
Damush et al., 2003a (Damush et al., 2003b)	AIMS2 symptoms subscale	no	Intervention:	6.2 (2.2)	77	4.7 (2.8)	76	0.07 [-0.23, 0.38]	11.25	3.8 (2.5)	63	0.28 [-0.05, 0.62]
			Control:	6 (2.2)	87	4.9 (2.6)	87			4.5 (2.4)	76	
Ferrari et al., 2005[d]	Variable: any pain? (4 levels)	no	Intervention:	-	-	1.94 (0.88)	54	-0.16 [-0.54, 0.21][e]	3	1.63 (0.93)	49	-0.13 [-0.52, 0.26]
			Control:	-	-	1.8 (0.83)[e]	55			1.51 (0.87)	53	
Gatchel et al., 2003	CPI	no	Intervention:	-	-	-	-	-	11.25	26.8 (NA)	22	NA
			Control:	-	-	-	-			43.1 (NA)	48	
Gatchel et al., 2006 (Stowell et al., 2007)	CPI	no	Intervention:	58.8 (11.8)	56	-	-	-	10.5	22.4 (17.5)	56	0.52 [0.12, 0.92]
			Control:	57.3 (12.5)	45	-	-			33.3 (24[f])	45	

Table C1 (*Continued*)

Study[a]	Measure	high value desirable?		baseline		post-treatment			follow-up			
				M (*SD*)	*n*	*M* (*SD*)	*n*	*g* [95%-CI]	length of follow-up (months)[b]	*M* (*SD*)	*n*	*g* [95%-CI]
Gil-Jardiné et al., 2018	11-point NRS (pain intensity)	no	Intervention: EMDR:	*Med* = 5.5 (*IQR*: 4 - 7)	34	*Med* = 3 (*IQR*: 0.25 - 5)	34	NA	-	-	-	-
			Intervention: Reassurance:	*Med* = 6 (*IQR*: 3 - 7)	38	*Med* = 5 (*IQR*: 0 - 6)	38			-	-	
			Control:	*Med* = 5 (*IQR*: 3-7)	37	*Med* = 4 (*IQR*: 0 - 7)	37			-	-	
Hazard et al., 2000	11-point self-assessment of pain	no	Intervention:	-	-	NA	NA	NA	6	NA	NA	NA
			Control:	-	-	NA	NA			NA	NA	
Irvine et al., 2015	Item: how bad is your LBP?	no	Intervention:	0.96 (1.26)	199	0.82 (1.22)	199	0.25 [0.05, 0.45]	2	0.56 (1)	199	0.34 [0.14, 0.54]
			Control:	1.09 (1.34)	199	1.16 (1.47)	199			0.98 (1.43)	199	
Janse et al., 2016	CIS fatigue severity subscale	no	Intervention:	46.8 (5.58)	50	32.8 (14.8)	50	0.67 [0.27, 1.07]	-	-	-	-
			Control:	46.6 (4.89)	50	41.6 (11.1)	50			-	-	
Karjalainen et al., 2004 (Karjalainen et al., 2003)[g]	11-point NRS (pain intensity)	no	Intervention:	5.82 (Range: 1 - 10)	107	3.82 (Range: 0- 10)	104	NA	24	3.35 (Range: 0 - 9)	103	NA
			Control:	5.7 (Range: 1 -10)	57	4.1 (Range: 0 - 9)	56			3.4 (Range: 0 - 9)	53	
Kongsted et al., 2008	11-point box scale (neck pain)	no	Intervention:	*Med* = 2 (*IQR*: 2 - 3)	119	*Med* = 0 (*IQR*: 0 - 1)	64	NA	12	*Med* = 0 (*IQR*: 0 - 1)	103	NA
			Control:	*Med* = 2 (*IQR*: 2 - 3)	63	*Med* = 1 (*IQR*: 0 - 2)	27			*Med* = 1 (*IQR*: 0 - 3)	55	
Linton & Andersson, 2000[h] (Linton & Nordin, 2006)	11-point NRS (average back pain in the past week)	no	Intervention:	4.85 (2.61[i])	107	-	-	-	60	3.82 (NA)	87	NA
			Control:	4.78 (1.89[i])	70	-	-			NA	59	
Linton & Ryberg, 2001	11-point NRS (mean pain in the past week)	no	Intervention:	5.41 (1.67)	75	-	-	-	12	4.53 (1.67)	75	0.08 [-0.23, 0.39]
			Control:	5.58 (1.78)	85	-	-			4.67 (1.78)	85	
Nyenhuis, Zastrutzki, Weise, et al., 2013 (Nyenhuis, Zastrutzki, Jäger, et al., 2013)[j]	TQ	no	Intervention:	37.3 (15.1)	227	21.6 (16.6)	150	0.34 [0.04, 0.65]	9	19.5 (14.4)	136	0.36 [0.03, 0.69]
			Control:	34.5 (13)	77	27.4 (18)	58			25.2 (19.1)	49	
Riddle et al., 2019[k]	WOMAC Pain Scale, 3.1 Likert version	no	Intervention:	11.6 (3.1)	130	-	-	-	10.5	3.3 (4.89)	130	-0.061 [-0.3, 0.18]
			Control:	11.3 (3.5)	135	-	-			3 (4.99)	135	
Sanders et al., 2013	CPI	no	Intervention:	63.7 (12.5)	90	45 (18.2)	90	0.12 [-0.18, 0.42]	-	-	-	-
			Control:	65 (12.6)	81	47.1 (18.2)	81			-	-	

Table C1 (*Continued*)

Study[a]	Measure	high value desirable?		baseline		post-treatment			follow-up			
				M (*SD*)	*n*	*M* (*SD*)	*n*	*g* [95%-CI]	length of follow-up (months)[b]	*M* (*SD*)	*n*	*g* [95%-CI]
Sharpe et al., 2012 (study 1)	VAS	no	Intervention:	-	-	27.7 (21.7)	27	0.32 [-0.22, 0.85]	3	6.44 (11.1)	23	0.89 [0.28, 1.49]
			Control:	-	-	34.8 (22.8)	27			19.3 (16.8)	23	
Silverberg et al., 2013	RPQ	no	Intervention:	33.5 (13.3)	15	-	-	-	1.5	17.9 (14.5)	13	0.72 [-0.11, 1.55]
			Control:	37.3 (13.4)	13	-	-			28.7 (14.5)	11	
Slater et al., 2009	DDS	no	Intervention:	11.5 (4.24)	34	NA	NA	NA	8.5	NA	NA	NA
			Control:	11.2 (4.39)	33	NA	NA			NA	NA	
Sterling et al., 2019	11-point NRS (pain in the past 24h)	no	Intervention:	5.4 (1.8)	53	2.5 (2.2)	51	0.52 [0.12, 0.91]	10.5	2.9 (2.3)	48	0.32 [-0.09, 0.72]
			Control:	5.2 (1.9)	54	3.7 (2.4)	50			3.7 (2.7)	46	
Toft et al., 2010[l]	SCL somatization subscale	no	Intervention:	*Med* = 23 (*IQR*: 20 - 30)	47	20.5 (6.9)	37	-0.24 [-1.18, 0.69][m]	24	18.1 (4.98)	38	0.31 [-0.62, 1.23][m]
			Control:	*Med* = 23 (*IQR*: 20 - 28)	63	19 (5.59)	50			20.1 (7.22)	54	
Traeger et al., 2019	11-point NRS (pain intensity in the past week)	no	Intervention:	6.3 (2.4)	101	3.2 (2.4)	98	-0.04 [-0.32, 0.24]	10.5	1.8 (2.2)	94	0.3 [0.01, 0.59]
			Control:	6.1 (2.2)	101	3.1 (2.2)	96			2.5 (2.4)	89	
Whitfill et al., 2010 (Rogerson et al., 2010)	CPI	no	Intervention:	5.23 (2.51)	58	NA	NA	NA	9.5	2.96 (2.82)	58	0.45 [0.05, 0.84]
			Control:	2.5 (2.45)	44	NA	NA			4.27 (3.01)	44	

Table C1 (*Continued*)

Study[a]	Measure	high value desirable?	baseline		post-treatment			follow-up			
			M (*SD*)	*n*	*M* (*SD*)	*n*	*g* [95%-CI]	length of follow-up (months)[b]	*M* (*SD*)	*n*	*g* [95%-CI]

Note. $g > 0$ indicates a better outcome in the intervention group. AIMS2: Arthritis Impact Measurement Scales. BPI: Brief Pain Inventory. BPI-SF: Brief Pain Inventory - Short Form. CIS: Checklist Individual Strength. CPI: Characteristic Pain Inventory. DDS: Descriptor Differential Scale. EMDR: Eye movement desensitization and reprocessing. *IQR*: Interquartile range. *Med*: Median. Minus (-): Not applicable to the respective study. LBP: Low back pain. MPI: Multidimensional Pain Inventory. NA: Missing data. NRS: Numeric rating scale. RPQ: Rivermead Postconcussion Symptoms Questionnaire. SCL: Symptom Check List. TQ: Tinnitus Questionnaire. VAS: Visual analogue scale. WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

[a] References in parentheses indicate duplicate reports.

[b] Length of follow-up is calculated in reference to the end of treatment. If length of follow-up varied between subjects within a study and no mean length of follow-up was reported, we selected the shortest possible length to obtain a conservative effect estimate.

[c] Post-treatment and follow-up mean values were derived from linear mixed models. Post-treatment and follow-up standard deviations were calculated from confidence intervals from linear mixed models using the *t*-distribution (Higgins & Deeks, 2008).

[d] Symptom severity was measured as categorical variable. For our analyses, we treated this variable as continuous. Post-treatment data were provided by the study authors.

[e] For this value, the authors provided us with data containing one observation too much. The authors informed us that this variable was derived from two items, one asking if the subject had no pain / was all better, and the other one assessing whether the subject currently had minor vs. moderate vs. severe pain. Therefore, the excess observation probably resulted from one subject answering both items. We decided for a conservative procedure by omitting one observation from the "minor pain" category. If one would omit an observation from the "no pain / was all better" category instead, this would result in a mean (SD) of 1.82 (0.8) and an effect of -0.14 [-0.52, 0.23]. Rerunning the analyses with these values did not change the pattern of findings.

[f] Reported values were inconsistent between reports. We extracted the value from the duplicate report since they seemed more realistic.

[g] Data from the mini-intervention and mini-intervention + work-site visit group were combined.

[h] Intervention: Cognitive behavior therapy group. Control: Pamphlet group.

[i] Calculated from reported confidence intervals using the *t*-distribution (Higgins & Deeks, 2008).

[j] Data extracted from completer analysis. Data from the Internet training, bibliotherapy and group treatment conditions were combined. The information-only group served as comparator.

[k] Follow-up mean values were derived from linear mixed models. Follow-up standard deviations were calculated from confidence intervals from linear mixed models using the *t*-distribution (Higgins & Deeks, 2008).

[l] Data extracted for subjects with sub-threshold somatoform disorder. Outcome data were provided by the study authors.

[m] The confidence interval of this effect size was computed via an imputed intracluster correlation coefficient (ICC) of 0.031, since we were not able to obtain the correct ICC for this outcome.

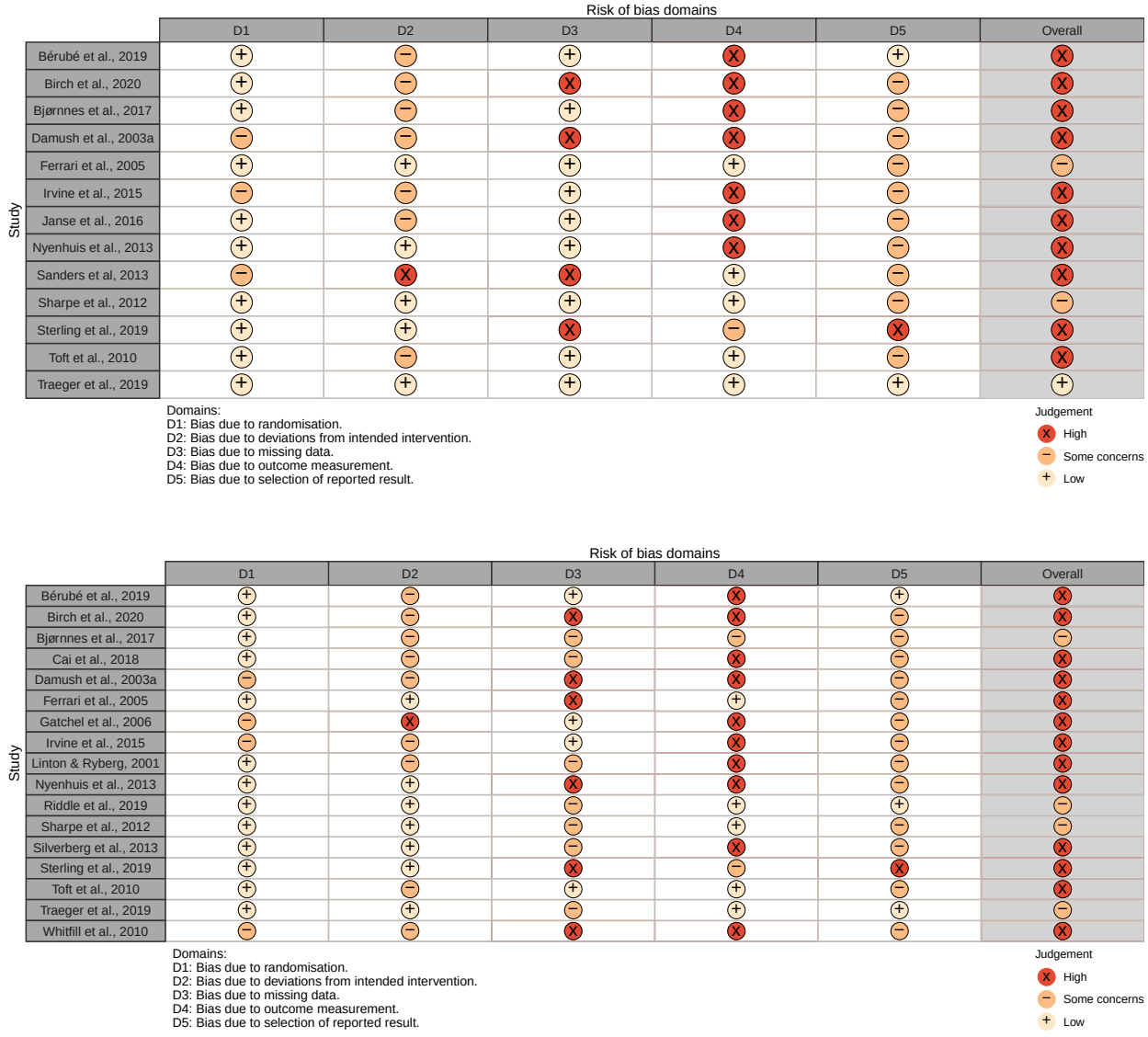

Figure C1. Risk of bias ratings for each study contributing somatic symptom severity effects. Upper panel: Post-treatment. Lower panel: Follow-up. The study by Toft et al. (2010) is a cluster-randomized trial. In this study, there was a high risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization (not depicted). Therefore, the overall risk of bias is rated as high for this study.

Table C2

Outcome data of studies measuring health-related quality of life

				baseline		post-treatment			follow-up			
Study[a]	Measure	high value desirable?		*M* (*SD*)	*n*	*M* (*SD*)	*n*	*g* [95%-CI]	length of follow-up (months)[b]	*M* (*SD*)	*n*	*g* [95%-CI]
Bérubé et al., 2019	BPI-I	no	Intervention:	7.3 (1.2)	28	2.7 (2.5)	25	0.06 [-0.5, 0.62	3	2.2 (2.5)	25	0 [-0.57, 0.57]
			Control:	7.5 (1.7)	28	2.85 (2.4)	24			2.2 (2.5)	23	
Birch et al., 2020[c]	EQ-5D	yes	Intervention:	0.58 (0.18)	29	0.72 (0.17)	24	-0.7 [-1.29, -0.11]	9	0.78 (0.19)	24	-0.5 [-1.07, 0.08]
			Control:	0.62 (2)	26	0.82 (0.1)	23			0.86 (0.12)	24	
Bjørnnes et al., 2017	BPI-I	no	Intervention:	-	-	NA	NA	NA	12	NA	NA	NA
			Control:	-	-	NA	NA			NA	NA	
Dahl & Nilsson, 2001	MPI interference	no	Intervention:	1.41 (1.14)	NA	-	-	-	12	1.09 (1.11)	NA	NA
	subscale		Control:	1.2 (0.78)	NA	-	-			0.92 (0.76)	NA	
Damush et al., 2003a	RDQ	no	Intervention:	14.7 (6.7)	77	11.7 (7.2)	76	0.11 [-0.2, 0.41]	11.25	9.1 (6.8)	63	0.29 [-0.05, 0.63]
(Damush et al., 2003b)			Control:	13.9 (6.8)	87	12.5 (7.7)	87			11.3 (8.1)	76	
Ferrari et al., 2005[d]	Item: limitation of	no	Intervention:	-	-	1.44 (0.9)	54	-0.13 [-0.51, 0.24]	3	0.82 (0.83)	49	0.06 [-0.33, 0.45]
	daily activities (4 levels)		Control:	-	-	1.31 (1.05)	55			0.87 (0.9)	53	
Irvine et al., 2015	Dartmouth CO-OP	no	Intervention:	20.4 (5.02)	199	19.3 (5.18)	199	0.28 [0.08, 0.47]	2	18.8 (5.39)	199	0.33 [0.13, 0.53]
			Control:	21 (4.96)	199	20.8 (5.92)	199			20.7 (5.64)	199	
Janse et al., 2016	SIP8	no	Intervention:	854 (478)	50	458 (577)	50	0.52 [0.12, 0.92]	-	-	-	-
			Control:	770 (411)	50	732 (455)	50			-	-	
Karjalainen et al., 2004	15D	yes	Intervention:	0.85 (Range: 0.61 - 1)	107	0.89 (Range: 0.6 - 1)	104	NA	24	0.89 (Range: 0.49 - 1)	103	NA
(Karjalainen et al., 2003)[e]			Control:	0.86 (Range: 0.7 - 0.98)	57	0.87 (Range: 0.6 - 1)	56			0.89 (Range: 0.6 - 1)	53	
Kongsted et al., 2008	Copenhagen Neck	no	Intervention:	-	-	*Med* = 2 (*IQR*: 0 - 5)	64	NA	12	*Med* = 1 (*IQR*: 0 - 3)	103	NA
	Functional Disability Scale		Control:	-	-	*Med* = 3 (*IQR*: 0 - 7)	27			*Med* = 2 (*IQR*: 0 - 3.25)	55	
Lamb et al., 2012	SF-12v1, PCS	yes	Intervention:	-	-	40.2 (8.9)	1771	-0.01 [-0.69, 0.67][g]	-	-	-	-
(Lamb et al., 2013)[f]			Control:	-	-	40.3 (9)	1305			-	-	
Linton & Andersson, 2000	EurQol 5	yes	Intervention:	-	-	-	-	-	60	NA	87	NA
(Linton & Nordin, 2006)[h]			Control:	-	-	-	-			NA	59	
Linton & Ryberg, 2001	6 physical functioning NRSs	yes	Intervention:	46 (11.2)	75	-	-	-	12	44.8 (11.6)	75	-0.04 [-0.35, 0.27]
			Control:	47.6 (10)	85	-	-			45.3 (11.4)	85	
Newcomer et al., 2008	ODI	no	Intervention:	25.2 (14.9)	69	NA	NA	NA	-	-	-	-
			Control:	26.1 (17.9)	69	NA	NA			-	-	

Table C2 (*Continued*)

Study[a]	Measure	high value desirable?		baseline		post-treatment			follow-up			
				M (*SD*)	*n*	*M* (*SD*)	*n*	*g* [95%-CI]	length of follow-up (months)[b]	*M* (*SD*)	*n*	*g* [95%-CI]
Riddle et al., 2019[i]	WOMAC Physical Function Scale	no	Intervention:	38.6 (11.8)	130	-	-	-	10.5	12.2 (18.4)	130	-0.027 [-0.27, 0.21]
			Control:	37.1 (11.8)	135	-	-			11.7 (18.5)	135	
Sanders et al, 2013	GCPS	no	Intervention:	2.21 (2.21)	90	1.63 (0.79)	90	-0.05 [-0.35, 0.25]	-	-	-	-
			Control:	2.17 (0.54)	81	1.59 (0.69)	81			-	-	
Sharpe et al., 2012 (study 1)	RMDQ	no	Intervention:	6.89 (6.23)	27	-	-	-	3	1.56 (2.9)	23	0.22 [-0.36, 0.8]
			Control:	7.53 (5.86)	27	-	-			2.48 (4.9)	23	
Slater et al., 2009	SIP	no	Intervention:	11.4 (7.69)	34	NA	NA	NA	8.5	NA	NA	NA
			Control:	12.3 (8.8)	33	NA	NA			NA	NA	
Sterling et al., 2019	NDI	no	Intervention:	44.9 (13.9)	53	25.5 (18.5)	51	0.43 [0.04, 0.82]	10.5	23.6 (20.2)	50	0.27 [-0.13, 0.67]
			Control:	41.7 (11.2)	55	33.1 (16.4)	51			28.7 (17.1)	48	
Toft et al., 2010[j]	SF-36, physical functioning subscale	yes	Intervention:	*Med* = 95 (*IQR*: 80 - 100)	45	92.8 (10.9)	38	0.27 [-0.66, 1.2][g]	24	92.2 (11.9)	38	0.27 [-0.65, 1.2][g]
			Control:	*Med* = 95 (*IQR*: 85 - 95)	57	89.5 (13)	51			87.8 (18.4)	54	
Traeger et al., 2019	RMDQ	no	Intervention:	11 (5.4)	101	5.6 (5.2)	98	0.27 [-0.01, 0.55]	10.5	3 (4.7)	94	0.16 [-0.13, 0.45]
			Control:	11.7 (5.8)	101	7.1 (5.8)	96			3.8 (5.1)	89	
Whitfill et al., 2010 (Rogerson et al., 2010)	SF-36	yes	Intervention:	33 (8.09)	58	NA	NA	NA	9.5	40.5 (11.5)	58	0.09 [-0.3, 0.48]
			Control:	36 (10.1)	44	NA	NA			39.5 (10.6)	44	

Note. $g > 0$ indicates a better outcome in the intervention group. BPI-I: Brief Pain Inventory - Interference subscales. Dartmouth CO-OP: Dartmouth Primary Care Cooperative Information Project scale. GCPS: Graded Chronic Pain Scale. *IQR*: Interquartile range. *Med*: Median. Minus (-): Not applicable to the respective study. MPI: Multidimensional Pain Inventory. NA: Missing data. NDI: Neck Disability Index. NRS: Numeric rating scale. ODI: Oswestry Disability Index. RDQ: Roland Disability Questionnaire. RMDQ: Roland-Morris Disability Questionnaire. SF-12v1, PCS: Short Form Questionnaire-12 Version 1, Physical Component Score. SF-36: Short Form Questionnaire-36. SIP: Sickness Impact Profile (Gilson et al., 1975). SIP8: Sickness Impact Profile (Bergner, Bobbitt, Carter, & Gilson, 1981). WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

[a] References in parentheses indicate duplicate reports.

[b] Length of follow-up is calculated in reference to the end of treatment. If length of follow-up varied between subjects within a study and no mean length of follow-up was reported, we selected the shortest possible length to obtain a conservative effect estimate.

[c] Post-treatment and follow-up mean values were derived from linear mixed models. Post-treatment and follow-up standard deviations were calculated from confidence intervals from linear mixed models using the *t*-distribution (Higgins & Deeks, 2008).

[d] Health-related quality of life was measured as categorical variable. For our analyses, we treated this variable as continuous. Post-treatment data were provided by the study authors.

[e] Data from the mini-intervention and mini-intervention + work-site visit group were combined.

[f] This study was a two-stepped randomized-controlled trial. No follow-up data were extracted since these data were contaminated by the second-step randomized-controlled trial evaluating physiotherapy.

[g] The confidence interval of this effect size was computed via an imputed intracluster correlation coefficient (ICC) of 0.031, since we were not able to obtain the correct ICC for this outcome.

[h] Intervention: Cognitive behavior therapy group. Control: Pamphlet group.

[i] Follow-up mean values were derived from linear mixed models. Follow-up standard deviations were calculated from confidence intervals from linear mixed models using the *t*-distribution (Higgins & Deeks, 2008).

[j] Data extracted for subjects with sub-threshold somatoform disorder. Outcome data were provided by the study authors.

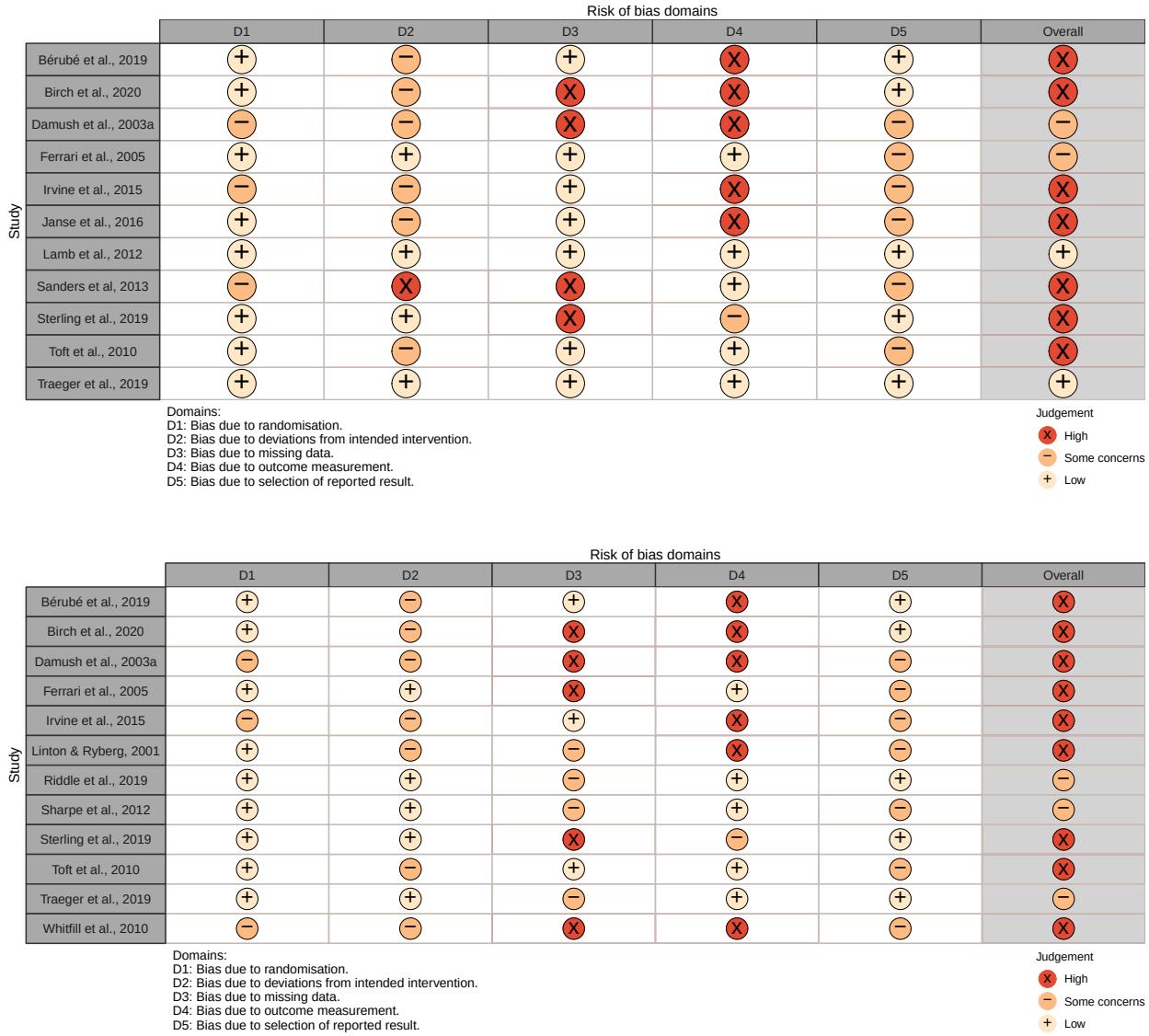

Figure C2. Risk of bias ratings for each study contributing health-related quality of life effects. Upper panel: Post-treatment. Lower panel: Follow-up. The studies by Lamb et al. (2012) and Toft et al. (2010) are cluster-randomized trials. While the study by Lamb et al. (2012) was at low risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization, the study by Toft et al. (2010) was at high risk (not depicted). Therefore, the overall risk of bias in the study by Toft et al. (2010) is rated as high.

Table C3

Outcome data of studies measuring diagnostic status concerning SSD/FSS

Study[a]	Measure		baseline		post-treatment			follow-up			
			yes | no	*n*	yes | no	*n*	*RR* [95%-CI]	length of follow-up (months)[b]	yes | no	*n*	*RR* [95%-CI]
Gatchel et al., 2006	SCID-I, pain disorder	Intervention:	NA	NA	-	-	-	10.5	NA	56	*OR* = 0.11 [0.04, 0.29][c]
(Stowell et al., 2007)		Control:	NA	NA	-	-			NA	45	
Gil-Jardiné et al., 2018[d]	standardized questionnaire administered by researcher	Intervention:	-	-	-	-	-	3	29 | 55	84	0.54 [0.37, 0.78]
	based on DSM-IV-TR diagnostic criteria for PCS	Control:	-	-	-	-			25 | 14	39	
Janse et al., 2016	combination of CIS fatigue, SIP8,	Intervention:	-	-	36 | 14	50	0.78 [0.65, 0.95]	-	-	-	-
	SF-36 physical and social functioning	Control:	-	-	46 | 4	50			-	-	
Kongsted et al., 2008	combination of pain and work status	Intervention:	-	-	44 | 20	64	1.24 [0.85, 1.8]	12	72 | 31	103	1.2 [0.93, 1.55]
		Control:	-	-	15 | 12	27			32 | 23	55	
Sanders et al, 2013	RDC/TMD	Intervention:	65 | 25	90	48 | 42	90	1.08 [0.81, 1.45]	-	-	-	-
		Control:	50 | 31	81	40 | 41	81			-	-	
Slater et al., 2009	combination of DDS and SIP	Intervention:	-	-	16 | 17	33	0.71 [0.46, 1.08]	12	NA	NA	NA
		Control:	-	-	22 | 10	32			NA	NA	

Note. *RR* < 1 indicates a better outcome in the intervention group. CIS: Checklist Individual Strength. DDS: Descriptor Differential Scale. Minus (-): Not applicable to the respective study. NA: Missing data. *OR*: Odds ratio. PCS: Post-concussion syndrome. RDC/TMD: Research Diagnostic Criteria for Temporomandibular Disorders. *RR*: Risk ratio. SCID-I: Structured Clinical Interview for DSM-IV Axis I Disorders. SF-36: Short Form Questionnaire-36. SIP: Sickness Impact Profile (Gilson et al., 1975). SIP8: Sickness Impact Profile (Bergner et al., 1981).

[a] References in parentheses indicate duplicate reports.

[b] Length of follow-up is calculated in reference to the end of treatment. If length of follow-up varied between subjects within a study and no mean length of follow-up was reported, we selected the shortest possible length to obtain a conservative effect estimate.

[c] Values converted from reported *OR* and its confidence interval in order to match our alignment of effect sizes (values < 1 indicating a better outcome in the intervention group).

[d] Data based on worst-case scenario analysis designating subjects abandoning the protocol after randomization for reasons related to clinical worsening or early discharge as having an SSD/FSS.

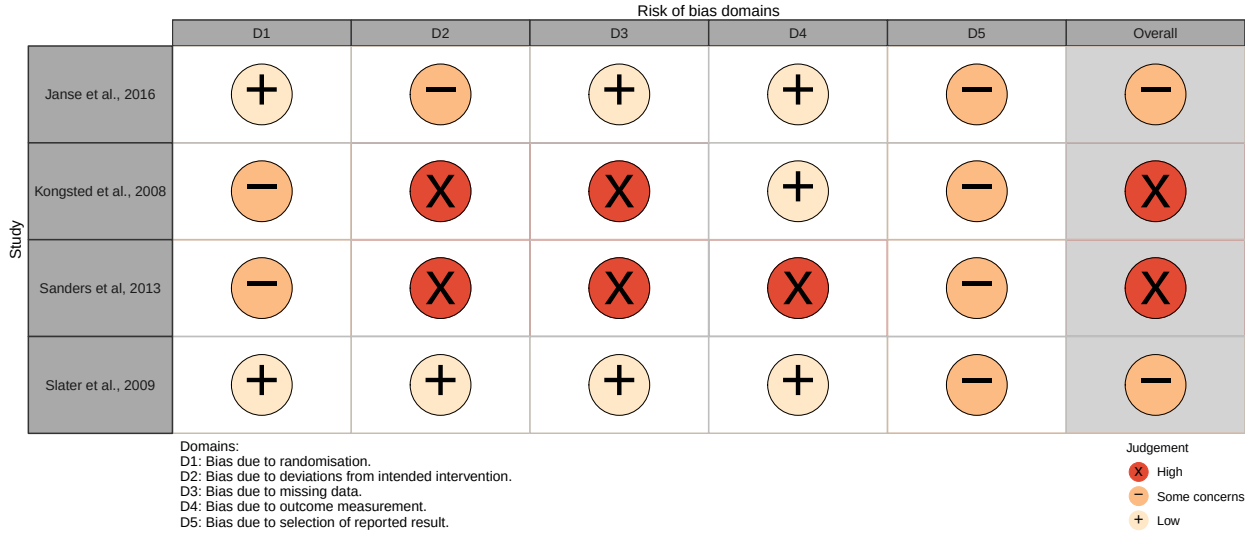

Figure C3. Risk of bias ratings for each study contributing diagnostic status concerning SSD/FSS effects post-treatment.

Table C4

Outcome data of studies measuring anxiety

Study[a]	Measure	Source	High value desirable?		baseline		post-treatment			follow-up			
					M (*SD*)	*n*	*M* (*SD*)	*n*	*g* [95%-CI]	length of follow-up (months)[b]	*M* (*SD*)	*n*	*g* [95%-CI]
Bérubé et al., 2019	HADS, anxiety subscale	self-report	no	Intervention:	8.3 (5)	28	6.4 (4.6)	25	-0.19 [-0.75, 0.37]	3	6.1 (5)	25	-0.022 [-0.59, 0.54]
				Control:	8.9 (4.1)	28	5.6 (3.5)	24			6 (3.7)	23	
Linton & Andersson, 2000 (Linton & Nordin, 2006)[c]	HADS, anxiety subscale	self-report	no	Intervention:	5.3 (3.65[d])	107	-	-	-	12	5.3 (3.62[d])	92	-0.06 [-0.38, 0.27]
				Control:	6.1 (4.61[d])	70	-	-			5.1 (3.57[d])	63	
Linton & Ryberg, 2001	HADS, anxiety subscale	self-report	no	Intervention:	6.17 (3.82)	75	-	-	-	12	5.5 (3.49)	75	0.17 [-0.14, 0.48]
				Control:	6.42 (4.24)	85	-	-			6.17 (4.26)	85	
Newcomer et al., 2008	STAI, state anxiety	self-report	no	Intervention:	32.9 (10.7)	69	NA	NA	NA	-	-	-	-
				Control:	33.6 (9.4)	69	NA	NA			-	-	
Sharpe et al., 2012 (study 1)	DASS, anxiety subscale	self-report	no	Intervention:	NA	NA	-	-	-	3	NA	23	NA
				Control:	NA	NA	-	-			NA	23	
Silverberg et al., 2013	HADS, anxiety subscale	self-report	no	Intervention:	NA	NA	-	-	-	1.5	8.5 (2.9)	13	-0.03 [-0.84, 0.77]
				Control:	NA	NA	-	-			8.4 (2.9)	11	
Sterling et al., 2019	DASS, anxiety subscale	self-report	no	Intervention:	12 (10.3)	53	7 (8.4)	51	0 [-0.39, 0.39]	10.5	9.3 (9.8)	48	-0.27 [-0.67, 0.14]
				Control:	7.6 (8.3)	54	7 (7.9)	50			7 (6.9)	46	
Toft et al., 2010[e]	Whiteley-7	self-report	no	Intervention:	*Med* = 13 (*IQR*: 10 - 18)	47	2.51 (2.94)	37	0.036 [-0.9, 0.97][f]	24	1.95 (3.64)	38	0.17 [-0.76, 1.09][f]
				Control:	*Med* = 12 (*IQR*: 10 - 15)	63	2.64 (3.94)	50			2.57 (3.78)	54	

Note. $g > 0$ indicates a better outcome in the intervention group. DASS: Depression, Anxiety and Stress Scale. HADS: Hospital Anxiety and Depression Scale. *IQR*: Interquartile range. *Med*: Median. Minus (-): Not applicable to the respective study. NA: Missing data. STAI: Spielberger's State-Trait Anxiety Inventory.

[a] References in parentheses indicate duplicate reports.

[b] Length of follow-up is calculated in reference to the end of treatment. If length of follow-up varied between subjects within a study and no mean length of follow-up was reported, we selected the shortest possible length to obtain a conservative effect estimate.

[c] Intervention: Cognitive behavior therapy group. Control: Pamphlet group.

[d] Calculated from reported confidence intervals using the *t*-distribution (Higgins & Deeks, 2008).

[e] Data extracted for subjects with sub-threshold somatoform disorder. Outcome data were provided by the study authors.

Table C4 (*Continued*)

[f] The confidence interval of this effect size was computed via an imputed intracluster correlation coefficient (ICC) of 0.031, since we were not able to obtain the correct ICC for this outcome.

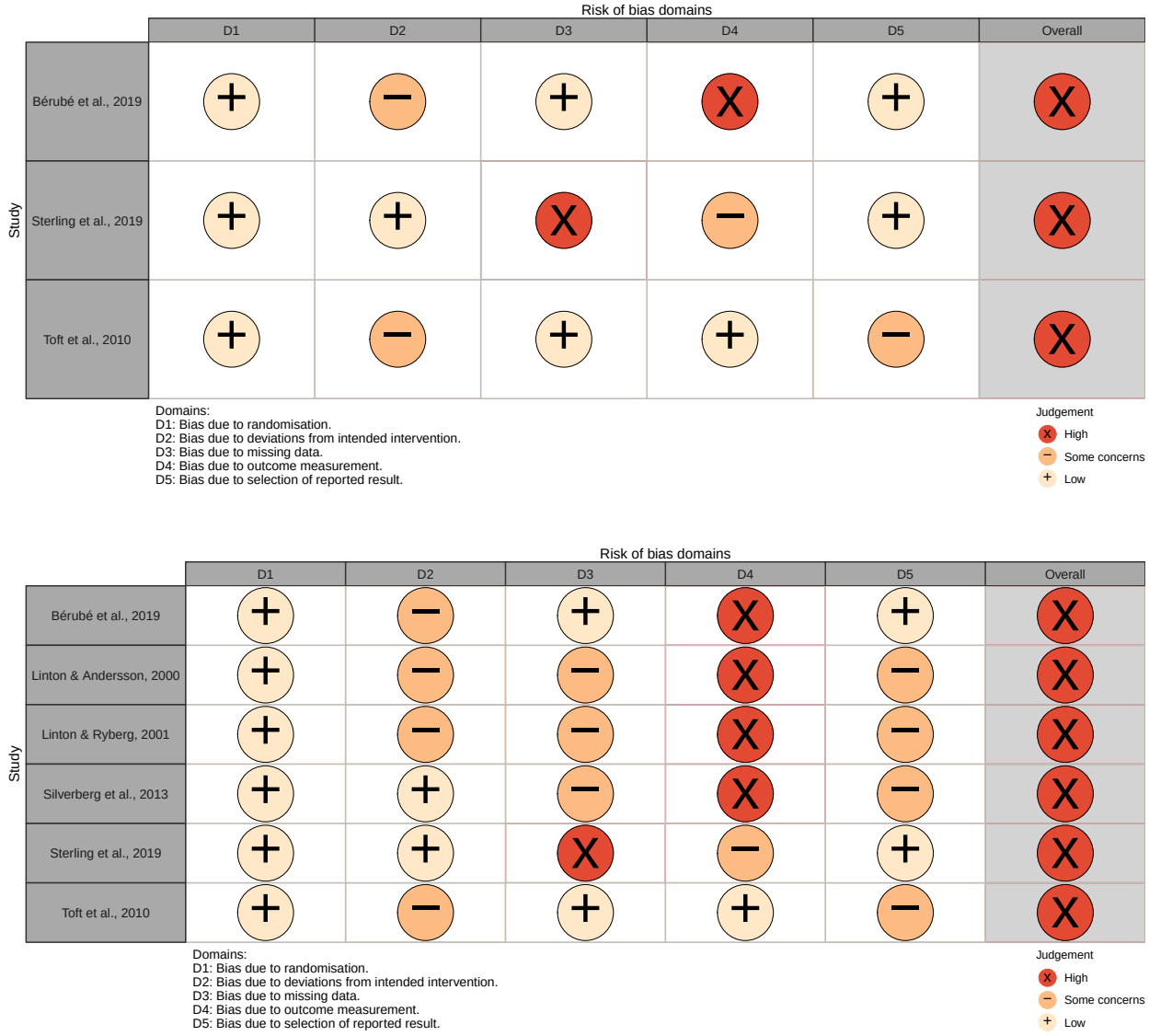

Figure C4. Risk of bias ratings for each study contributing anxiety effects at follow-up. Upper panel: Post-treatment. Lower panel: Follow-up. The study by Toft et al. (2010) is a cluster-randomized trial. In this study, there was a high risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization (not depicted). Therefore, the overall risk of bias is rated as high for this study.

Table C5

Outcome data of studies measuring depression

Study[a]	Measure	Source	High value desirable?		baseline		post-treatment			follow-up			
					M (*SD*)	*n*	*M* (*SD*)	*n*	*g* [95%-CI]	length of follow-up (months)[b]	*M* (*SD*)	*n*	*g* [95%-CI]
Bérubé et al., 2019	HADS, depression subscale	self-report	no	Intervention:	7.6 (4.6)	28	4 (3.4)	25	0.057 [-0.5, 0.62]	3	3.3 (3.6)	25	0.2 [-0.36, 0.77]
				Control:	8.4 (4.8)	28	4.2 (3.5)	24			4.1 (4.1)	23	
Gatchel et al., 2006	BDI-II	self-report	no	Intervention:	9.52 (9.95)	56	-	-	-	10.5	5.27 (6.98)	56	0.34 [-0.06, 0.73]
(Stowell et al., 2007)				Control:	8.71 (6.82)	45	-	-			7.98 (9)	45	
Linton & Andersson, 2000	HADS, depression subscale	self-report	no	Intervention:	4.2 (3.65[d])	107	-	-	-	12	3.9 (3.62[d])	92	0 [-0.32, 0.32]
(Linton & Nordin, 2006)[c]				Control:	4 (3.36[d])	70	-	-			3.9 (2.98[d])	63	
Linton & Ryberg, 2001	HADS, depression subscale	self-report	no	Intervention:	4.84 (3.46)	75	-	-	-	12	4.41 (3.77)	75	0.04 [-0.27, 0.35]
				Control:	4.74 (3.29)	85	-	-			4.56 (3.78)	85	
Nyenhuis, Zastrutzki, Weise, et al., 2013	PHQ-D, 9-item short form	self-report	no	Intervention:	7.27 (5.07)	227	5.42 (5.12)	150	0.06 [-0.25, 0.36]	9	5.72 (4.82)	136	-0.004 [-0.33, 0.32]
(Nyenhuis, Zastrutzki, Jäger, et al., 2013)[e]				Control:	5.9 (4.6)	77	5.7 (4.8)	58			5.7 (5.1)	49	
Sanders et al, 2013	BDI-II	self-report	no	Intervention:	8.73 (6.94)	88	5.65 (6.94)	88	0.07 [-0.23, 0.38]	-	-	-	-
				Control:	7.96 (6.98)	80	6.16 (6.98)	80			-	-	
Sharpe et al., 2012 (study 1)	DASS, depression subscale	self-report	no	Intervention:	NA	NA	-	-	-	3	NA	23	NA
				Control:	NA	NA	-	-			NA	23	
Silverberg et al., 2013	HADS, depression subscale	self-report	no	Intervention:	NA	NA	-	-	-	1.5	5 (3.1)	13	0.72 [-0.11, 1.54]
				Control:	NA	NA	-	-			7.3 (3.1)	11	
Sterling et al., 2019	DASS, depression subscale	self-report	no	Intervention:	12.7 (10.7)	53	7.8 (8.8)	51	0.22 [-0.18, 0.61]	10.5	9.7 (10.5)	48	0 [-0.4, 0.4]
				Control:	9.4 (9.3)	54	9.7 (8.7)	50			9.7 (9.9)	46	
Traeger et al., 2019	DASS, depression subscale	self-report	no	Intervention:	4.1 (3.7)	101	2.6 (4.1)	98	0.17 [-0.12, 0.45]	1.5[f]	2.1 (3.9)	97	0.1 [-0.18, 0.38]
				Control:	5.1 (5)	101	3.3 (4.3)	96			2.5 (4.1)	97	
Whitfill et al., 2010	BDI	self-report	no	Intervention:	11.6 (9.3)	58	NA	NA	NA	9.5	8.81 (9.49)	58	0.13 [-0.26, 0.52]
(Rogerson et al., 2010)				Control:	9.43 (9.58)	44	NA	NA			10.1 (10.2)	44	

Note. $g > 0$ indicates a better outcome in the intervention group. BDI: Beck Depression Inventory. BDI-II: Beck Depression Inventory (2nd Ed.). DASS: Depression, Anxiety and Stress Scale. HADS: Hospital Anxiety and Depression Scale. Minus (-): Not applicable to the respective study. NA: Missing data. PHQ-D: Patient Health Questionnaire.

[a] References in parentheses indicate duplicate reports.

[b] Length of follow-up is calculated in reference to the end of treatment. If length of follow-up varied between subjects within a study and no mean length of follow-up was reported, we selected the shortest possible length to obtain a conservative effect estimate.

[c] Intervention: Cognitive behavior therapy group. Control: Pamphlet group.

[d] Calculated from reported confidence intervals using the *t*-distribution (Higgins & Deeks, 2008).

Table C5 (*Continued*)

[e] Data extracted from completer analysis. Data from the Internet training, bibliotherapy and group treatment conditions were combined. The information-only group served as comparator.

[f] The corresponding author clarified that this was the last measurement of depression in this study.

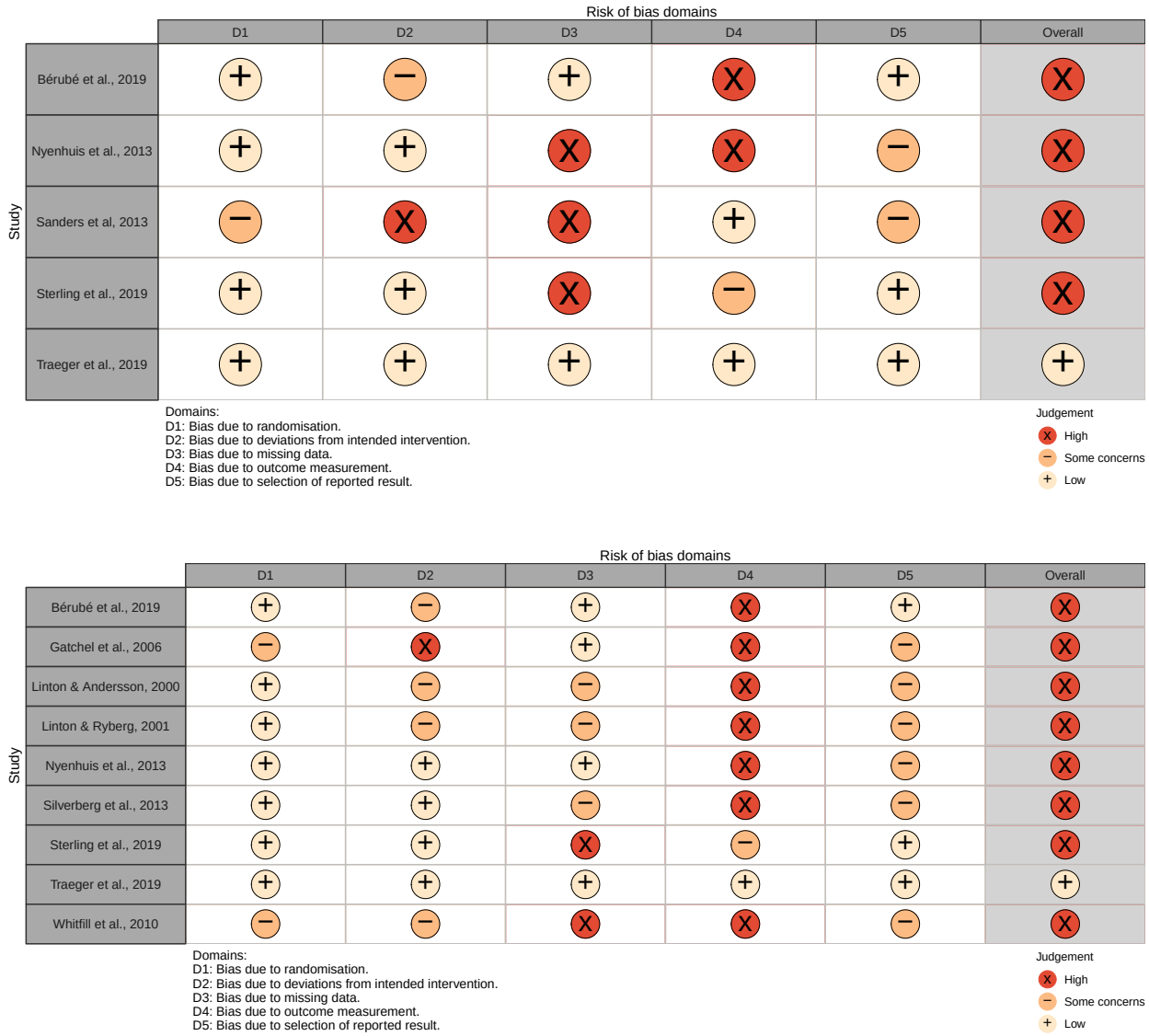

Figure C5. Risk of bias ratings for each study contributing depression effect data. Upper panel: Post-treatment. Lower panel: Follow-up.

Table C6

Outcome data of studies measuring health care utilization

Study[a]	Measure	Source	High value desirable?		baseline		post-treatment			follow-up			
					M (*SD*)	*n*	*M* (*SD*)	*n*	*g* [95%-CI]	length of follow-up (months)[b]	*M* (*SD*)	*n*	*g* [95%-CI]
Gatchel et al., 2003	health care visits related to LBP	self-report	no	Intervention:	-	-	-	-	-	11.25	17 (NA)	22	NA
				Control:	-	-	-	-			27.3 (NA)	48	
Gatchel et al., 2006	visits to health care providers	self-report	no	Intervention:	NA	NA	-	-	-	10.5	1.67 (5.46)	54	0.32 [-0.08, 0.71]
(Stowell et al., 2007)	during the study period			Control:	NA	NA	-	-			4.09 (9.54)	45	
Hazard et al., 2000	health care visits	self-report	no	Intervention:	-	-	NA	NA	NA	6	NA	NA	NA
				Control:	-	-	NA	NA			NA	NA	
Karjalainen et al., 2004	visits to a physician during past 3 months	self-report	no	Intervention:	3.56 (Range: 0 - 20)	107	NA	104	NA	24	NA	103	NA
(Karjalainen et al., 2003)[c]				Control:	3.3 (Range: 0 - 18)	57	NA	56			NA	53	
Linton & Andersson, 2000	visits to a physician, physical therapist,	self-report	no	Intervention:	5.78 (NA)	107	-	-	-	60	5.97 (NA)	87	NA
(Linton & Nordin, 2006)[d]	specialist or hospital, and alternative			Control:	7 (NA)	70	-	-			NA	59	
	care provider during past year												
Linton & Ryberg, 2001	visits to a physician, physical therapist,	self-report	no	Intervention:	3.69 (5.45)	75	-	-	-	12	4.51 (6.25)	75	0.29 [-0.03, 0.6]
	specialist or hospital, and alternative			Control:	5.25 (7.48)	85	-	-			6.62 (8.19)	85	
	care provider during past year												
Newcomer et al., 2008	LBP-related physician office	objective data	no	Intervention:	-	-	NA	NA	NA	-	-	-	-
	visits in the past year			Control:	-	-	NA	NA			-	-	
Nyenhuis, Zastrutzki, Weise, et al., 2013 [e]	doctor's appointments related	self-report	no	Intervention:	NA	NA	NA	NA	NA	9	NA	NA	NA
(Nyenhuis, Zastrutzki, Jäger, et al., 2013)	to tinnitus during past 4 weeks			Control:	NA	NA	NA	NA			NA	NA	
Silverberg et al., 2013	family physician visits	self-report	no	Intervention:	-	-	-	-	-	1.5	5.7 (3.9)	13	0.45 [-0.37, 1.26]
	during the trial period			Control:	-	-	-	-			7.8 (5.2)	11	

Note. $g > 0$ indicates a better outcome in the intervention group. LBP: Low back pain. Minus (-): Not applicable to the respective study. NA: Missing data.

[a] References in parentheses indicate duplicate reports.

[b] Length of follow-up is calculated in reference to the end of treatment. If length of follow-up varied between subjects within a study and no mean length of follow-up was reported, we selected the shortest possible length to obtain a conservative effect estimate.

[c] Data from the mini-intervention and mini-intervention + work-site visit group were combined.

[d] Intervention: Cognitive behavior therapy group. Control: Pamphlet group.

[e] Data from the Internet training, bibliotherapy and group treatment conditions were combined. The information-only group served as comparator.

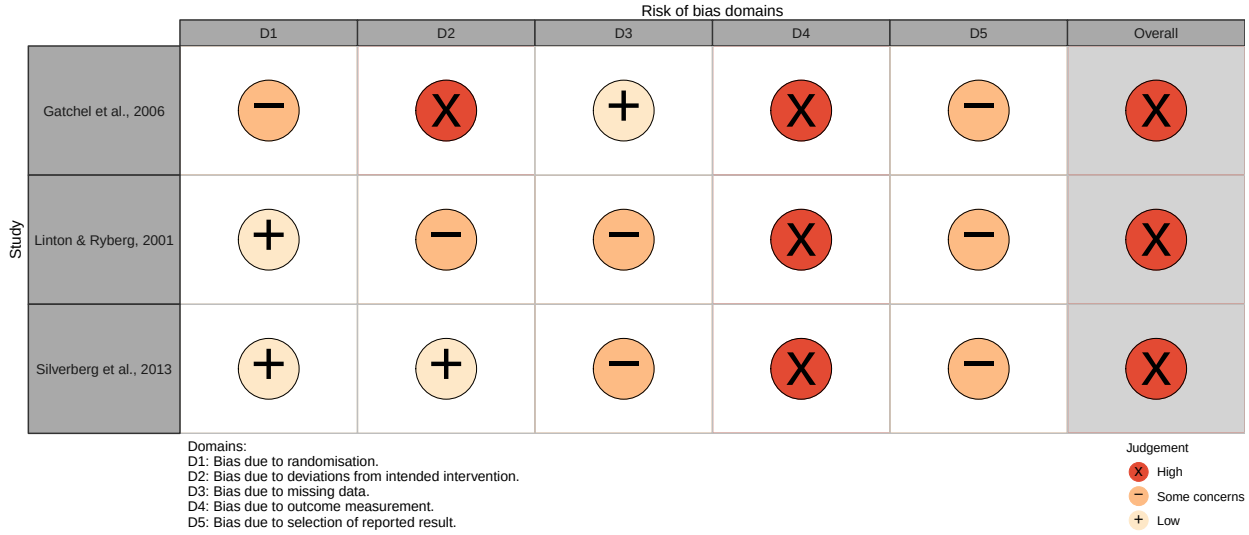

Figure C6. Risk of bias ratings for each study contributing health care utilization effects at follow-up.

Table C7

Outcome data of studies measuring consumer satisfaction

Study[a]	Measure	High value desirable?		baseline		post-treatment			follow-up			
				M (*SD*)	*n*	*M* (*SD*)	*n*	*g* [95%-CI]	length of follow-up (months)[b]	*M* (*SD*)	*n*	*g* [95%-CI]
Damush et al., 2003a	LBP treatment	yes	Intervention:	24.4 (5)	77	25.5 (5.3)	76	-0.02 [-0.33, 0.29]	11.25	25.9 (4.6)	63	0.1 [-0.24, 0.43]
(Damush et al., 2003b)	satisfaction scale		Control:	24.3 (5.7)	87	25.6 (5.4)	87			25.4 (5.4)	76	
Gil-Jardiné et al., 2018	11-point NRS	yes	Intervention:			EMDR:	EMDR:	NA	-			-
				-	-	*Med* = 9.5 (*IQR*: 8 - 10)	34			-	-	
						Reassurance:	Reassurance:					
				-	-	*Med* = 8.5 (*IQR*: 7.25 - 10)	38			-	-	
			Control:	-	-	*Med* = 8 (*IQR*: 6 - 10)	37			-	-	
Karjalainen et al., 2004	11-point NRS	yes	Intervention:	4.36 (Range: 0 - 9)	107	6.15 (Range: 0- 10)	104	NA	24	5.99 (Range: 0 - 10)	103	NA
(Karjalainen et al., 2003)[c]			Control:	4.1 (Range: 0 - 10)	57	4.1 (Range: 0 - 10)	56			4.3 (Range: 0 - 10)	53	
Lamb et al., 2012	NA	NA	Intervention:	-	-	NA	NA	NA	-	-	-	-
(Lamb et al., 2013)			Control:	-	-	NA	NA			-	-	
Linton & Andersson, 2000	Item: to what extent did	yes	Intervention:	-	-	-	-	-	12	NA	92	NA
(Linton & Nordin, 2006)[d]	you find this intervention		Control:	-	-	-	-			NA	63	
	to be of help? (5 levels)											
Nyenhuis, Zastrutzki, Weise, et al., 2013 [e]	11-point rating scale	yes	Intervention:	-	-	7.4 (2.64)	150	1.21 [0.89, 1.54]	-	-	-	-
(Nyenhuis, Zastrutzki, Jäger, et al., 2013)			Control:	-	-	3.9 (3.4)	58			-	-	
Silverberg et al., 2013	5-point Likert scale	yes	Intervention:	-	-	-	-	-	1.5	4.69 (0.48)	13	NA
			Control:	-	-	-	-			NA	NA	
Slater et al., 2009	Item: would you	yes	Intervention:	-	-	NA	NA	NA	8.5	NA	NA	NA
	recommend the		Control:	-	-	NA	NA			NA	NA	
	treatment to others?											

Note. $g > 0$ indicates a better outcome in the intervention group. EMDR: Eye movement desensitization and reprocessing. *IQR*: Interquartile range. LBP: Low back pain. *Med*: Median. Minus (-): Not applicable to the respective study. NA: Missing data. NRS: Numeric rating scale.

[a] References in parentheses indicate duplicate reports.

[b] Length of follow-up is calculated in reference to the end of treatment. If length of follow-up varied between subjects within a study and no mean length of follow-up was reported, we selected the shortest possible length to obtain a conservative effect estimate.

[c] Data from the mini-intervention and mini-intervention + work-site visit group were combined.

[d] Intervention: Cognitive behavior therapy group. Control: Pamphlet group.

[e] Data from the Internet training, bibliotherapy and group treatment conditions were combined. The information-only group served as comparator.

Online Supplement D

Search results

Table D1

Number of literature search results by search term set and electronic database

Search term set	PubMed	PsycINFO	Web of Science
General terms	521	373	665
multisomatoform disorder	0	0	0
somatization disorder	26	37	2
pain disorder	1	7	1
conversion disorder	403	89	154
somatic symptom disorder	0	3	1
bodily distress disorder	1	0	0
bodily stress syndrome	0	0	0
neurasthenia	0	2	0
culture-bound syndrome	0	20	0
Allergology			
food intolerance	53	4	236
multiple chemical sensitivity	2	2	1
sick building syndrome	2	0	4
Persian gulf syndrome	0	0	1
amalgam hypersensitivity	1	0	1
implant intolerance	0	0	0
prosthesis intolerance	0	0	0
aerotoxic syndrome	0	0	0
Anesthesiology			
idiopathic pain	2	1	0
chronic postoperative pain	0	2	11
Cardiology			
atypical chest pain	7	2	3

Table D1 (*Continued*)

Search term set	PubMed	PsycINFO	Web of Science
palpitations with normal investigations	0	0	0
syndrome X	6	0	5
Dermatology			
psychogenic skin disease	0	0	0
Endocrinology			
hypoglycaemia	0	0	0
Gastroenterology	3	4	6
functional bowel disorders	2	1	0
irritable bowel syndrome	51	19	51
nonulcer dyspepsia	6	1	4
functional abdominal pain	0	0	0
functional colonical disease	0	0	0
functional disorders of swallowing	2	0	0
Gynecology			
premenstrual syndrome	9	7	4
Infectiology			
chronic lyme disease	0	2	75
candida hypersensitivity	0	0	0
chronic rhinopharyngitis	0	0	0
Neurology	0	0	2
functional seizures	0	3	5
functional voice disorder	0	1	2
functional motor/ movement/ sensorimotor disorder	46	21	7
functional eye movement disorder	0	0	0
functional facial/tongue movement disorder	0	0	0

Table D1 (*Continued*)

Search term set	PubMed	PsycINFO	Web of Science
functional sensory symptoms	0	1	1
functional visual symptoms	1	0	0
functional auditory disorders	41	18	33
functional speech disorder	2	1	0
functional memory/cognitive disorder	1	3	2
functional dizziness	295	119	21
functional stroke	5	0	1
tension headache	29	25	37
atypical face pain	20	3	1
electromagnetic hypersensitivity	1	1	0
central sensitivity syndrome	0	1	1
post-concussion syndrome	88	45	26
Oral medicine / Otorhinolaryngology			
temporomandibular joint disorder	28	10	42
atypical odontalgia	0	0	0
psychogenic gagging	0	0	0
burning mouth	2	0	2
bruxism	3	2	4
globus syndrome	0	0	0
Orthopedics			
repetitive strain injury	23	15	38
chronic whiplash syndrome	24	2	20
neck pain	77	28	45
Respiratory medicine			
hyperventilation syndrome	0	2	1
Rheumatology	0	0	0

Table D1 (*Continued*)

Search term set	PubMed	PsycINFO	Web of Science
fibromyalgia	69	55	51
chronic low back pain	396	84	481
chronic pain / persistent pain / chronic intractable benign pain syndrome	270	255	210
chronic fatigue syndrome / myalgic encephalomyelitis / post-viral fatigue syndrome	132	163	121
Urology			
functional urologic disorders	0	0	1
Fowler's syndrome	0	0	0
paruresis	0	0	0
dysfunctional voiding	0	0	2
idiopathic overactive bladder	0	0	0
interstitial cystitis	5	1	4
urethral syndrome	3	0	3
chronic pelvic pain syndrome	37	7	9
pelvic arthropathy	0	0	0
Total	2696	1442	2398

Note. In some medical specialties, there are umbrella terms for specialty-specific SSD/FSS. The number of search results for these specialty-specific umbrella terms are listed in the same row as the corresponding specialty heading.

Online Supplement E

List of included studies

Bérubé, M., Gélinas, C., Feeley, N., Martorella, G., Côté, J., Laflamme, G. Y., … Choinière, M. (2019). Feasibility of a hybrid web-based and in-person self-management intervention aimed at preventing acute to chronic pain transition after major lower extremity trauma (iPACT-E-Trauma): A pilot randomized controlled trial. *Pain Medicine*, *20*(10), 2018–2032. doi:10.1093/pm/pnz008

Birch, S., Stilling, M., Mechlenburg, I., & Hansen, T. B. (2020). No effect of cognitive behavioral patient education for patients with pain catastrophizing before total knee arthroplasty: A randomized controlled trial. *Acta Orthopaedica*, *91*(1), 98–103. doi:10.1080/17453674.2019.1694312

Bjørnnes, A. K., Parry, M., Lie, I., Fagerland, M. W., Watt-Watson, J., Rustøen, T., … Leegaard, M. (2017). The impact of an educational pain management booklet intervention on postoperative pain control after cardiac surgery. *European Journal of Cardiovascular Nursing*, *16*(1), 18–27. doi:10.1177/1474515116631680

Cai, L., Gao, H., Xu, H., Wang, Y., Lyu, P., & Liu, Y. (2018). Does a program based on cognitive behavioral therapy affect kinesiophobia in patients following total knee arthroplasty? A randomized, controlled trial with a 6-month follow-up. *The Journal of Arthroplasty*, *33*(3), 704–710. doi:10.1016/j.arth.2017.10.035

Dahl, J. C., & Nilsson, A. (2001). Evaluation of a randomized preventive behavioural medicine work site intervention for public health workers at risk for developing chronic pain. *European Journal of Pain*, *5*(4), 421–432. doi:10.1053/eujp.2001.0264

Damush, T. M., Weinberger, M., Perkins, S. M., Rao, J. K., Tierney, W. M., Qi, R., & Clark, D. O. (2003a). Randomized trial of a self-management program for primary care patients with acute low back pain: Short-term effects. *Arthritis and Rheumatism*, *49*(2), 179–186. doi:10.1002/art.10995

Damush, T. M., Weinberger, M., Perkins, S. M., Rao, J. K., Tierney, W. M., Qi, R., & Clark, D. O. (2003b). The long-term effects of a self-management program for

inner-city primary care patients with acute low back pain. *Archives of Internal Medicine*, *163*(21), 2632–2638. doi:10.1001/archinte.163.21.2632

Ferrari, R., Rowe, B. H., Majumdar, S. R., Cassidy, J. D., Blitz, S., Wright, S. C., & Russell, A. S. (2005). Simple educational intervention to improve the recovery from acute whiplash: Results of a randomized, controlled trial. *Academic Emergency Medicine*, *12*(8), 699–706. doi:10.1197/j.aem.2005.03.531

Gatchel, R. J., Polatin, P. B., Noe, C., Gardea, M., Pulliam, C., & Thompson, J. (2003). Treatment- and cost-effectiveness of early intervention for acute low-back pain patients: A one-year prospective study. *Journal of Occupational Rehabilitation*, *13*(1), 1–9. doi:10.1023/a:1021823505774

Gatchel, R. J., Stowell, A. W., Wildenstein, L., Riggs, R., & Ellis, E. (2006). Efficacy of an early intervention for patients with acute temporomandibular disorder–related pain: A one-year outcome study. *The Journal of the American Dental Association*, *137*(3), 339–347. doi:10.14219/jada.archive.2006.0183

Gil-Jardiné, C., Evrard, G., Al Joboory, S., Tortes Saint Jammes, J., Masson, F., Ribéreau-Gayon, R., … Lagarde, E. (2018). Emergency room intervention to prevent post concussion-like symptoms and post-traumatic stress disorder. A pilot randomized controlled study of a brief eye movement desensitization and reprocessing intervention versus reassurance or usual care. *Journal of Psychiatric Research*, *103*, 229–236. doi:10.1016/j.jpsychires.2018.05.024

Hazard, R. G., Reid, S., Haugh, L. D., & McFarlane, G. (2000). A controlled trial of an educational pamphlet to prevent disability after occupational low back injury. *Spine*, *25*(11), 1419–1423. doi:10.1097/00007632-200006010-00015

Irvine, A. B., Russell, H., Manocchia, M., Mino, D. E., Cox Glassen, T., Morgan, R., … Ary, D. V. (2015). Mobile-web app to self-manage low back pain: Randomized controlled trial. *Journal of Medical Internet Research*, *17*(1), e1, 1–21. doi:10.2196/jmir.3130

Janse, A., Wiborg, J. F., Bleijenberg, G., Tummers, M., & Knoop, H. (2016). The efficacy of guided self-instruction for patients with idiopathic chronic fatigue: A randomized

controlled trial. *Journal of Consulting and Clinical Psychology*, *84*(5), 377–388. doi:10.1037/ccp0000085

Karjalainen, K., Malmivaara, A., Mutanen, P., Roine, R., Hurri, H., & Pohjolainen, T. (2004). Mini-intervention for subacute low back pain: Two-year follow-up and modifiers of effectiveness. *Spine*, *29*(10), 1069–1076. doi:10.1097/00007632-200405150-00004

Karjalainen, K., Malmivaara, A., Pohjolainen, T., Hurri, H., Mutanen, P., Rissanen, P., … Roine, R. (2003). Mini-intervention for subacute low back pain: A randomized controlled trial. *Spine*, *28*(6), 533–540. doi:10.1097/01.BRS.0000049928.52520.69

Kongsted, A., Qerama, E., Kasch, H., Bach, F. W., Korsholm, L., Jensen, T. S., & Bendix, T. (2008). Education of patients after whiplash injury: Is oral advice any better than a pamphlet? *Spine*, *33*(22), E843–848. doi:10.1097/BRS.0b013e318182bee2

Lamb, S. E., Gates, S., Williams, M. A., Williamson, E. M., Mt-Isa, S., Withers, E. J., … Managing Injuries of the Neck Trial (MINT) Study Team. (2013). Emergency department treatments and physiotherapy for acute whiplash: A pragmatic, two-step, randomised controlled trial. *Lancet*, *381*(9866), 546–556. doi:10.1016/S0140-6736(12)61304-X

Lamb, S. E., Williams, M. A., Williamson, E. M., Gates, S., Withers, E. J., Mt-Isa, S., … MINT Trial Group. (2012). Managing injuries of the neck trial (MINT): A randomised controlled trial of treatments for whiplash injuries. *Health Technology Assessment*, *16*(49), 1–141. doi:10.3310/hta16490

Linton, S. J., & Andersson, T. (2000). Can chronic disability be prevented? A randomized trial of a cognitive-behavior intervention and two forms of information for patients with spinal pain. *Spine*, *25*(21), 2825–2831. doi:10.1097/00007632-200011010-00017

Linton, S. J., & Nordin, E. (2006). A 5-year follow-up evaluation of the health and economic consequences of an early cognitive behavioral intervention for back pain: A randomized, controlled trial. *Spine*, *31*(8), 853–858. doi:10.1097/01.brs.0000209258.42037.02

Linton, S. J., & Ryberg, M. (2001). A cognitive-behavioral group intervention as prevention for persistent neck and back pain in a non-patient population: A randomized controlled trial. *Pain*, *90*(1), 83–90. doi:10.1016/s0304-3959(00)00390-0

Mitchell, R. I., & Carmen, G. M. (1994). The functional restoration approach to the treatment of chronic pain in patients with soft tissue and back injuries. *Spine*, *19*(6), 633–642. doi:10.1097/00007632-199403001-00001

Newcomer, K. L., Vickers Douglas, K. S., Shelerud, R. A., Long, K. H., & Crawford, B. (2008). Is a videotape to change beliefs and behaviors superior to a standard videotape in acute low back pain? A randomized controlled trial. *The Spine Journal*, *8*(6), 940–947. doi:10.1016/j.spinee.2007.08.007

Nyenhuis, N., Zastrutzki, S., Jäger, B., & Kröner-Herwig, B. (2013). An internet-based cognitive-behavioural training for acute tinnitus: Secondary analysis of acceptance in terms of satisfaction, trial attrition and non-usage attrition. *Cognitive Behaviour Therapy*, *42*(2), 139–145. doi:10.1080/16506073.2012.724081

Nyenhuis, N., Zastrutzki, S., Weise, C., Jäger, B., & Kröner-Herwig, B. (2013). The efficacy of minimal contact interventions for acute tinnitus: A randomised controlled study. *Cognitive Behaviour Therapy*, *42*(2), 127–138. doi:10.1080/16506073.2012.655305

Riddle, D. L., Keefe, F. J., Ang, D. C., Slover, J., Jensen, M. P., Bair, M. J., … Dumenci, L. (2019). Pain coping skills training for patients who catastrophize about pain prior to knee arthroplasty: A multisite randomized clinical trial. *JBJS*, *101*(3), 218–227. doi:10.2106/JBJS.18.00621

Rogerson, M. D., Gatchel, R. J., & Bierner, S. M. (2010). A cost utility analysis of interdisciplinary early intervention versus treatment as usual for high-risk acute low back pain patients. *Pain Practice*, *10*(5), 382–395. doi:10.1111/j.1533-2500.2009.00344.x

Sanders, C., Liegey-Dougall, A., Lorduy, K., Haggard, R., & Gatchel, R. J. (2013). The effects of an early intervention program on physical symptoms in an acute temporomandibular disorder population: A preliminary study. *Journal of Applied Biobehavioral Research*, *18*(4), 218–230. doi:10.1111/jabr.12011

Sharpe, L., Ianiello, M., Dear, B. F., Nicholson Perry, K., Refshauge, K., & Nicholas, M. K. (2012). Is there a potential role for attention bias modification in pain patients? Results

of 2 randomised, controlled trials. *Pain*, *153*(3), 722–731. doi:10.1016/j.pain.2011.12.014

Silverberg, N. D., Hallam, B. J., Rose, A., Underwood, H., Whitfield, K., Thornton, A. E., & Whittal, M. L. (2013). Cognitive-behavioral prevention of postconcussion syndrome in at-risk patients: A pilot randomized controlled trial. *The Journal of Head Trauma Rehabilitation*, *28*(4), 313–322. doi:10.1097/HTR.0b013e3182915cb5

Slater, M. A., Weickgenant, A. L., Greenberg, M. A., Wahlgren, D. R., Williams, R. A., Carter, C., … Atkinson, J. H. (2009). Preventing progression to chronicity in first onset, subacute low back pain: An exploratory study. *Archives of Physical Medicine and Rehabilitation*, *90*(4), 545–552. doi:10.1016/j.apmr.2008.10.032

Sterling, M., Smeets, R., Keijzers, G., Warren, J., & Kenardy, J. (2019). Physiotherapist-delivered stress inoculation training integrated with exercise versus physiotherapy exercise alone for acute whiplash-associated disorder (StressModex): A randomised controlled trial of a combined psychological/physical intervention. *British Journal of Sports Medicine*, *53*(19), 1240–1247. doi:10.1136/bjsports-2018-100139

Stowell, A. W., Gatchel, R. J., & Wildenstein, L. (2007). Cost-effectiveness of treatments for temporomandibular disorders: Biopsychosocial intervention versus treatment as usual. *Journal of the American Dental Association*, *138*(2), 202–208. doi:10.14219/jada.archive.2007.0137

Toft, T., Rosendal, M., Ørnbøl, E., Olesen, F., Frostholm, L., & Fink, P. (2010). Training general practitioners in the treatment of functional somatic symptoms: Effects on patient health in a cluster-randomised controlled trial (the functional illness in primary care study). *Psychotherapy and Psychosomatics*, *79*(4), 227–237. doi:10.1159/000313691

Traeger, A. C., Lee, H., Hübscher, M., Skinner, I. W., Moseley, G. L., Nicholas, M. K., … McAuley, J. H. (2019). Effect of intensive patient education vs placebo patient education on outcomes in patients with acute low back pain: A randomized clinical trial. *JAMA Neurology*, *76*(2), 161–169. doi:10.1001/jamaneurol.2018.3376

Whitfill, T., Haggard, R., Bierner, S. M., Pransky, G., Hassett, R. G., & Gatchel, R. J. (2010). Early intervention options for acute low back pain patients: A randomized clinical trial

with one-year follow-up outcomes. *Journal of Occupational Rehabilitation*, *20*(2), 256–263. doi:10.1007/s10926-010-9238-4

Online Supplement F

Systematic review and meta-analysis of secondary outcomes

Secondary outcomes

Unwanted negative treatment effects.

Post-treatment. Since only one study assessed unwanted negative treatment effects post-treatment, we describe these data narratively. Sterling, Smeets, Keijzers, Warren, Kenardy (2019) evaluated the effect of stress inoculation training in combination with guideline-based exercise compared to guideline-based exercise alone (SC/TAU) for patients suffering from whiplash-associated disorder (n = 108). The researchers assessed adverse effects (i.e., exacerbation of a pre-existing condition) and adverse events (i.e., events that are life-threatening, require inpatient hospitalization, or will result in persistent or significant disability or incapacity) via open ended questions. In each trial arm, one subject reported neck pain exacerbation, while no subject reported adverse events.

Follow-up. Only two studies assessed unwanted negative treatment effects at follow-up. Therefore, we describe these data narratively. In the study by Traeger et al. (2019), patients with acute low back pain (n = 202) were randomized to an intensive patient education condition or to a placebo education condition. Both treatments were delivered face-to-face. The researchers recorded adverse events during the trial. Over a follow-up time of 10.5 months, there were no reported adverse events in any of the treatment groups.

In the study by Riddle et al. (2019), patients scheduled for a knee arthroplasty at risk for chronic pain (n = 402) received either CBT-based pain coping skills training or arthritis education serving as placebo condition. Beyond that, there was a third trial arm providing SC/TAU, only. Unwanted negative treatment effects were assessed during data collection and by medical record review after a follow-up time of 10.5 months. There were no significant differences neither in adverse events (e.g., emergency room visits due to knee pain, psychological distress, elevated depressive symptoms) nor in serious adverse events (e.g., hospitalization, surgery, infection, death) between groups.

Diagnostic status concerning SSD/FSS. Outcome data for studies measuring diagnostic status concerning SSD/FSS can be found in Table E3.

Post-treatment. Four studies measured diagnostic status concerning SSD/FSS post-treatment. Effect size data were available for all of them (n = 427). A random-effects meta-analysis revealed a risk ratio of 0.92 (95%-CI: [0.62, 1.37], see Figure F1). Heterogeneity was not significantly different from zero ($Q(3) = 7.53$, $p = .057$) and inconsistency was small to considerable ($I^2 = 59.8\%$, 95%-CI: [0%, 97.5%]). The resulting 95%-prediction interval ranged from 0.45 to 1.89.

Risk of bias in individual studies. Figure F2 depicts the risk of bias inherent in the diagnostic status summary effect. See Figure E3 for risk of bias ratings for each study.

Meta-bias. The PET-PEESE revealed a corrected risk ratio of 0.68 (95%-CI: [0.12, 3.83]). The 3PSM revealed a corrected risk ratio of 0.96 (95%-CI: [0.66, 1.42]). A likelihood-ratio test did not reveal a significantly better fit of 3PSM to the data ($\chi^2(1) = 0.4$, p = .53).

Follow-up. Out of three studies measuring diagnostic status concerning SSD/FSS at follow-up, appropriate effect size data were available for two studies. Therefore, these data are

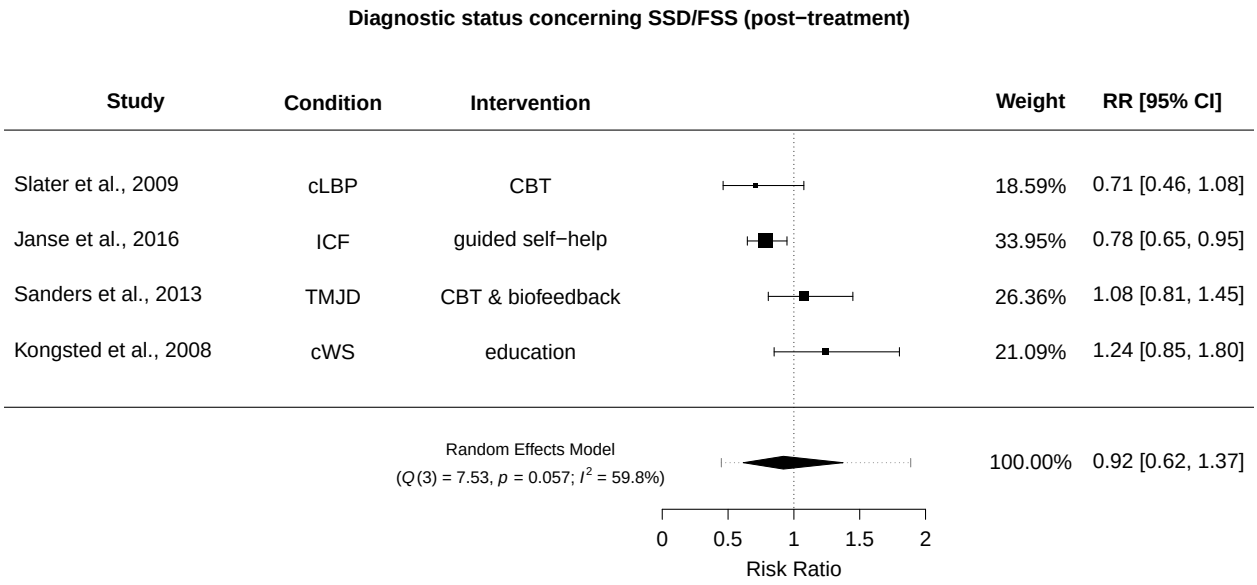

Figure F1. Forest plot of diagnostic status concerning SSD/FSS (post-treatment). *RR* < 1 indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cWS: Chronic whiplash syndrome. ICF: Idiopathic chronic fatigue. TMJD: Temporomandibular joint disorder.

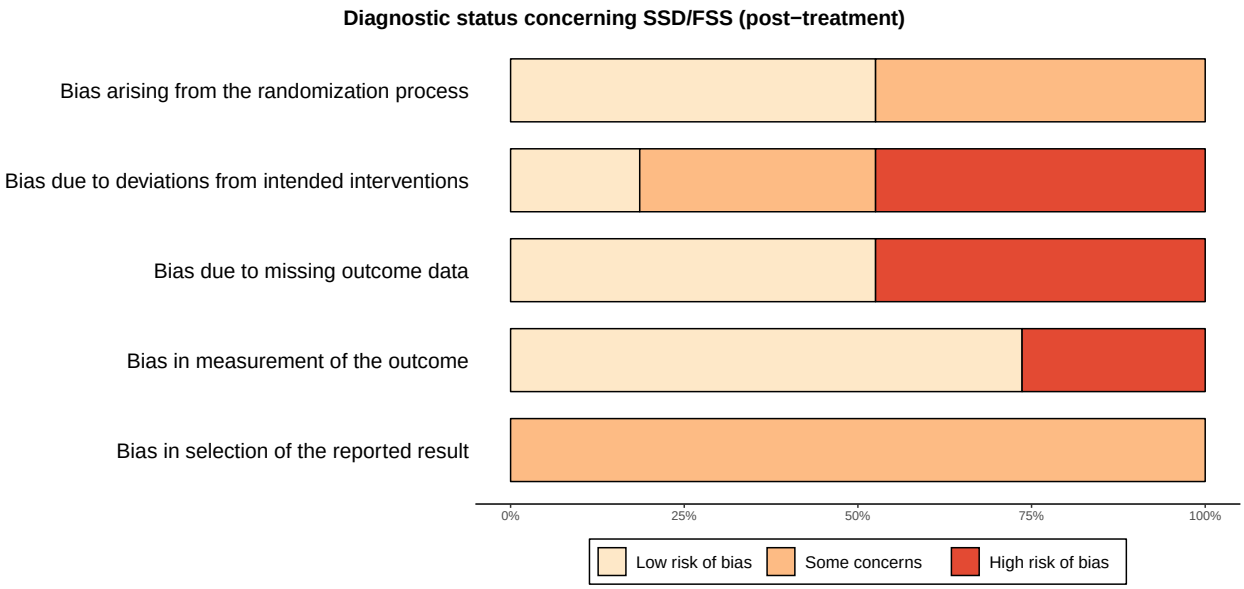

Figure F2. Risk of bias inherent in the summary effect for diagnostic status concerning SSD/FSS (post-treatment). Study-level biases are weighted according to the meta-analytic weights.

synthesized narratively. In the study by Gil-Jardiné et al. (2018), patients at high risk for developing a postconcussion syndrome were treated with a session of either EMDR or reassurance by a therapist in the emergency room. Control subjects received SC/TAU. Diagnostic status was determined via an interview based on the DSM-IV criteria for postconcussion syndrome. Based on a sample of n = 123 and a follow-up length of 3 months, there was a significant effect favoring the intervention groups (RR = 0.54, 95%-CI: [0.37, 0.78]). It is important to note that this effect stems from a worst-case-scenario analysis in which subjects abandoning the intervention protocol due to early discharge or clinical worsening were considered as having an SSD/FSS at follow-up.

Kongsted et al. (2008) examined the effect of oral advice given by a nurse at a home visit to patients presenting with a whiplash injury compared to SC/TAU consisting of an educational pamphlet. These patients were of comparably lower risk for chronic whiplash syndrome since patients at high risk were invited to participate in another trial. Diagnosis was defined via a combination of a neck pain measure and current work status. Based on a sample of 158 subjects and a follow-up length of 12 months, there was no significant effect of the

intervention (*RR* = 1.2, 95%-CI: [0.93, 1.55]).

Although the study by Gatchel, Stowell, Wildenstein, Riggs, and Ellis (2006) did not provide appropriate effect size data for meta-analytic integration, it reports the effect of the intervention in another effect size metric. Therefore, we describe this study here, too. The study evaluated a combined CBT and biofeedback treatment program for patients suffering from acute jaw pain at high risk for developing a temporomandibular joint disorder. Patients in the control group received no intervention in the context of the trial. Diagnosis was determined by fulfilling the criteria for a pain disorder using the Structured Clinical Interview for DSM-IV. Based on a sample of n = 101 and a follow-up length of 10.5 months, there was a significant positive effect of the intervention (odds ratio = 0.11, 95%-CI: [0.04, 0.29]).

Anxiety. Outcome data for studies measuring anxiety are summarized in Table E4.

Post-treatment. Out of four studies measuring anxiety post-treatment, effect size data were available for three of them (n = 237). There was a small and non-significant negative effect ($g = -0.052$, 95%-CI: [-0.33, 0.22], see Figure F3). Heterogeneity was not significantly different from zero ($Q(2) = 0.34$, $p = .84$) and inconsistency was small to considerable ($I^2 = 0\%$, 95%-CI: [0%, 84.6%]). The resulting 95%-prediction interval ranged from -0.33 to 0.22.

Risk of bias in individual studies. The risk of bias ratings for each domain are depicted in Figure F4. See Figure E4 for risk of bias ratings for each study.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = -0.018$ (95%-CI: [-3.58, 3.54]). No corrected effect estimate could be computed via 3PSM due to convergence problems.

Follow-up. Out of seven studies measuring anxiety at follow-up, effect size data were available for six studies (n = 573). Follow-up length ranged from 1.5 months to 24 months (*Median* = 11.25). There was a small and non-significant negative effect ($g = -0.01$, 95%-CI: [-0.19, 0.17], see Figure F5). Heterogeneity was not significantly different from zero ($Q(5)$ = 3.06, $p = .69$) and inconsistency was small to substantial ($I^2 = 0\%$, 95%-CI: [0%, 65.1%]). The resulting 95%-prediction interval ranged from -0.19 to 0.17.

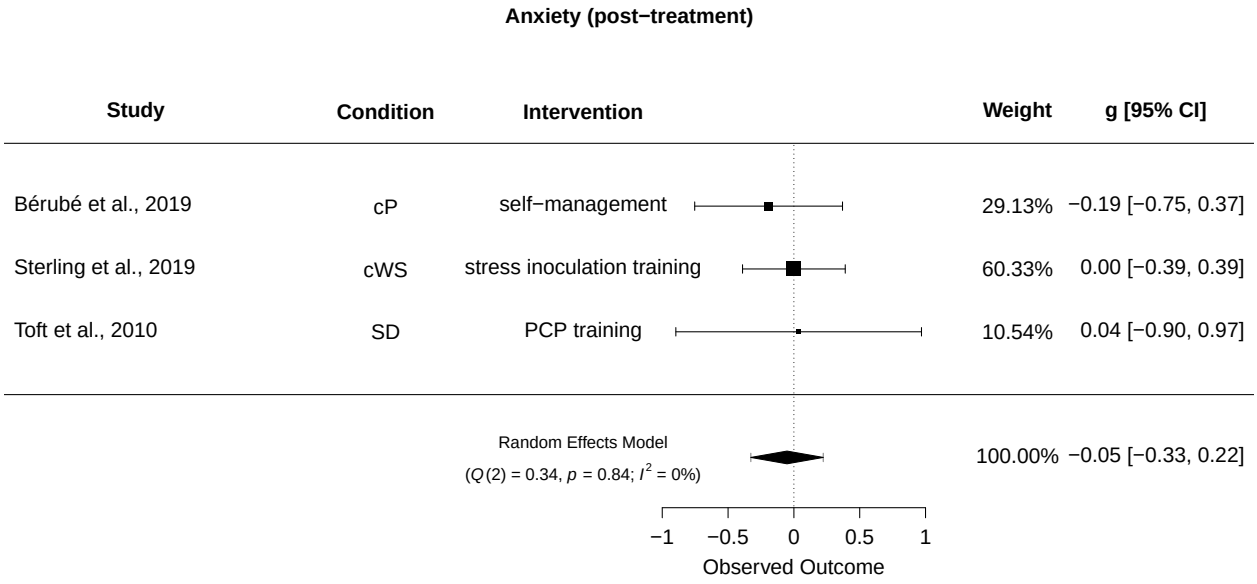

Figure F3. Forest plot of anxiety (post-treatment). $g > 0$ indicates more favorable outcomes in the intervention group. cP: Chronic pain. cWS: Chronic whiplash syndrome. PCP: Primary care physician. SD: Somatoform disorder.

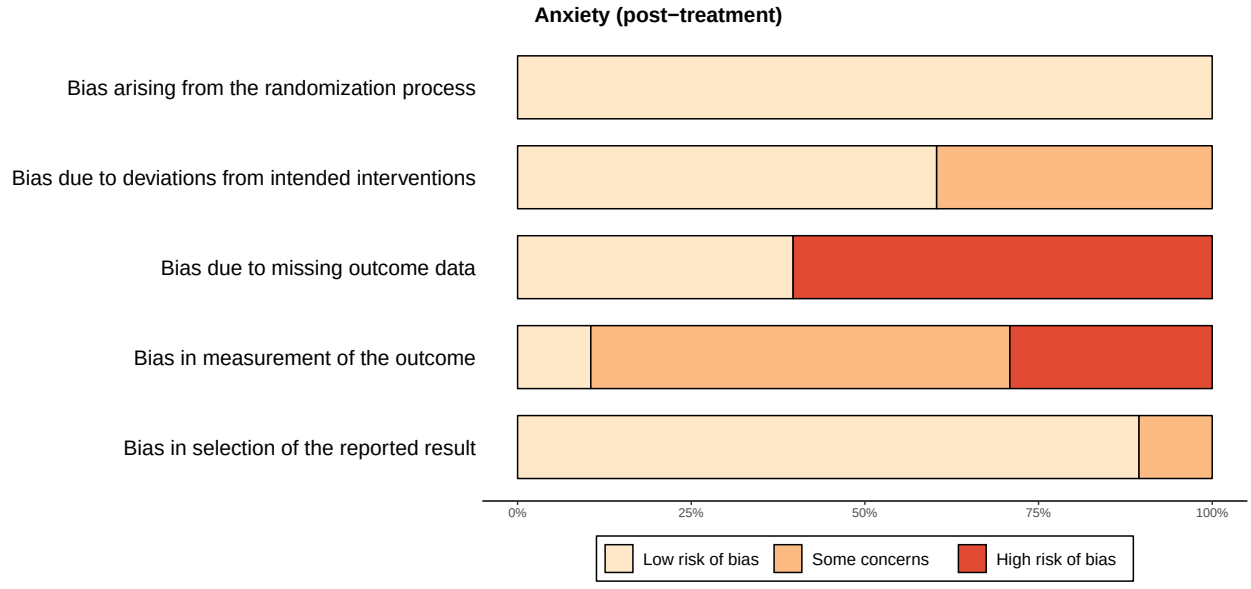

Figure F4. Risk of bias inherent in the summary effect for anxiety (post-treatment). Study-level biases are weighted according to the meta-analytic weights. One cluster-randomized study was included in this meta-analysis (Toft et al., 2010). There was a high risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization in this study (not depicted).

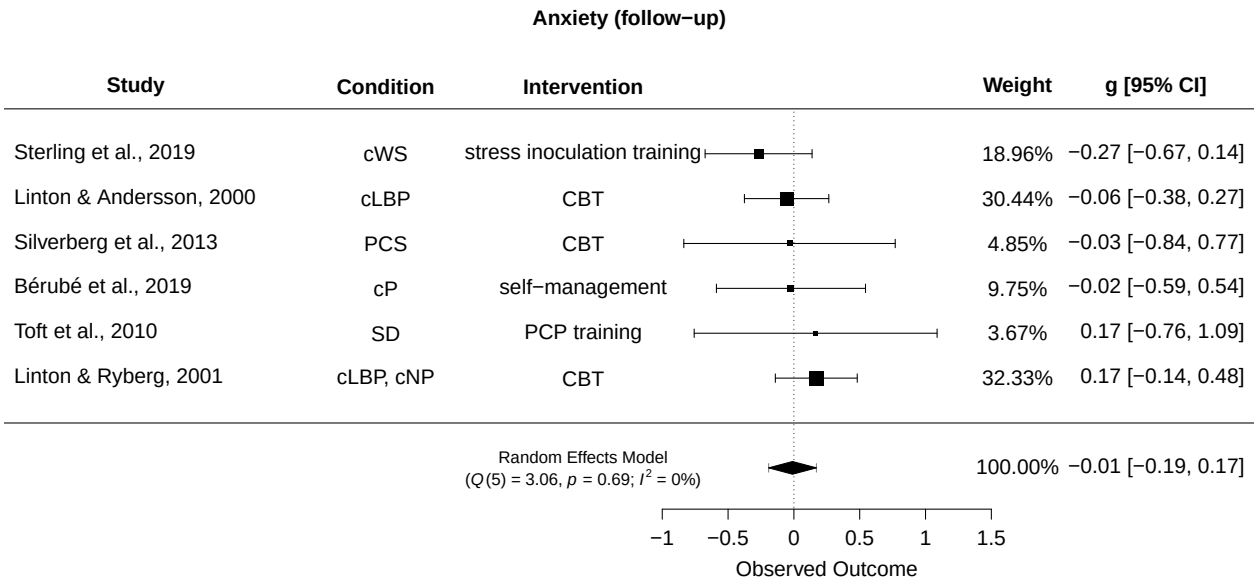

Figure F5. Forest plot of anxiety (follow-up). $g > 0$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. cP: Chronic pain. cWS: Chronic whiplash syndrome. PCP: Primary care physician. PCS: Post-concussion syndrome. SD: Somatoform disorder.

Risk of bias in individual studies. The risk of bias ratings for each domain are depicted in Figure F6. See Figure E4 for risk of bias ratings for each study.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.017$ (95%-CI: [-0.58, 0.61]). The 3PSM revealed a corrected effect estimate of $g = 0.003$ (95%-CI: [-0.19, 0.19]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 0.29$, p = .59).

Depression. Outcome data for studies measuring depression are listed in Table E5.

Post-treatment. Out of six studies measuring depression post-treatment, effect size data were available for five studies (n = 720). There was a small significant effect ($g = 0.12$, 95%-CI: [0.03, 0.2], see Figure F7). Heterogeneity was not significantly different from zero ($Q(4)$ = 0.64, $p = .96$) and inconsistency was small ($I^2 = 0\%$, 95%-CI: [0%, 24%]). The resulting 95%-prediction interval ranged from 0.03 to 0.2.

Risk of bias in individual studies. For a summary of risk of bias ratings, see Figure F8. See Figure E5 for risk of bias ratings for each study.

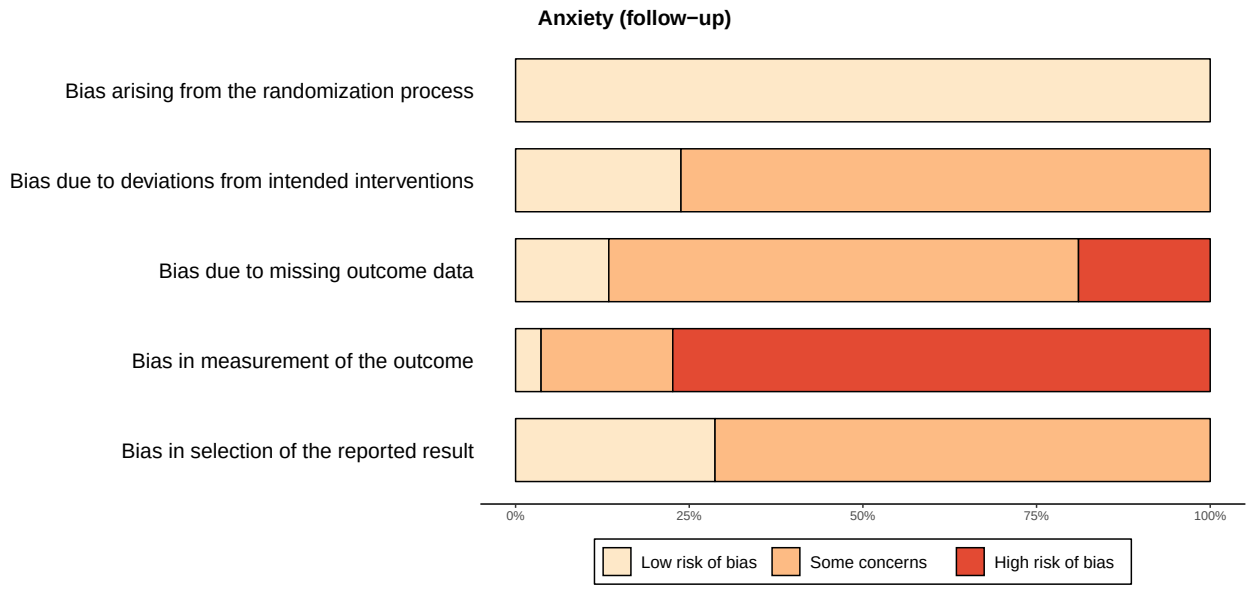

Figure F6. Risk of bias inherent in the summary effect for anxiety (follow-up). Study-level biases are weighted according to the meta-analytic weights. One cluster-randomized study was included in this meta-analysis (Toft et al., 2010). There was a high risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization in this study (not depicted).

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.12$ (95%-CI: [-0.4, 0.64]). The 3PSM revealed a corrected effect estimate of $g = 0.17$ (95%-CI: [0.046, 0.29]. A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 1.32$, $p = .25$).

Follow-up. Out of 10 studies measuring depression at follow-up, effect size data were available for nine studies (n = 1063). Follow-up length ranged from 1.5 months to 12 months (*Median* = 9.5). There was a small and non-significant effect ($g = 0.096$, 95%-CI: [-0.016, 0.21], see Figure F9). Heterogeneity was not significantly different from zero ($Q(8) = 4.83$, $p = .78$) and inconsistency was small to substantial ($I^2 = 0.015\%$, 95%-CI: [0%, 70.5%]). The resulting 95%-prediction interval ranged from -0.017 to 0.21.

Risk of bias in individual studies. Figure F10 depicts the risk of bias ratings. See Figure E5 for risk of bias ratings for each study.

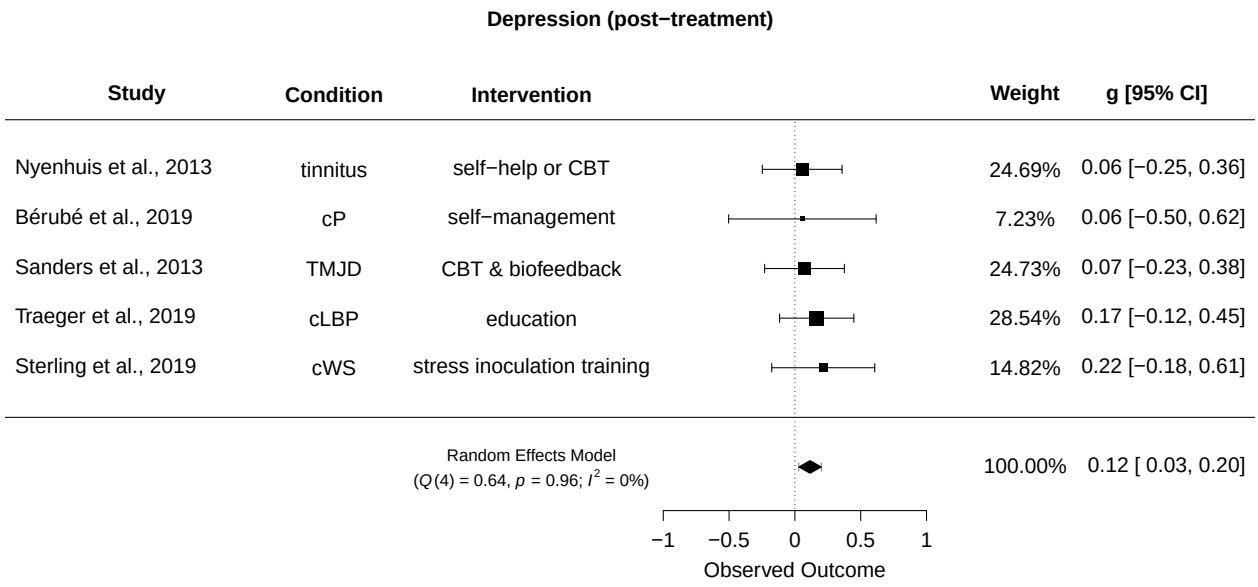

Figure F7. Forest plot of depression (post-treatment). $g > 1$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cP: Chronic pain. cWS: Chronic whiplash syndrome. TMJD: Temporomandibular joint disorder.

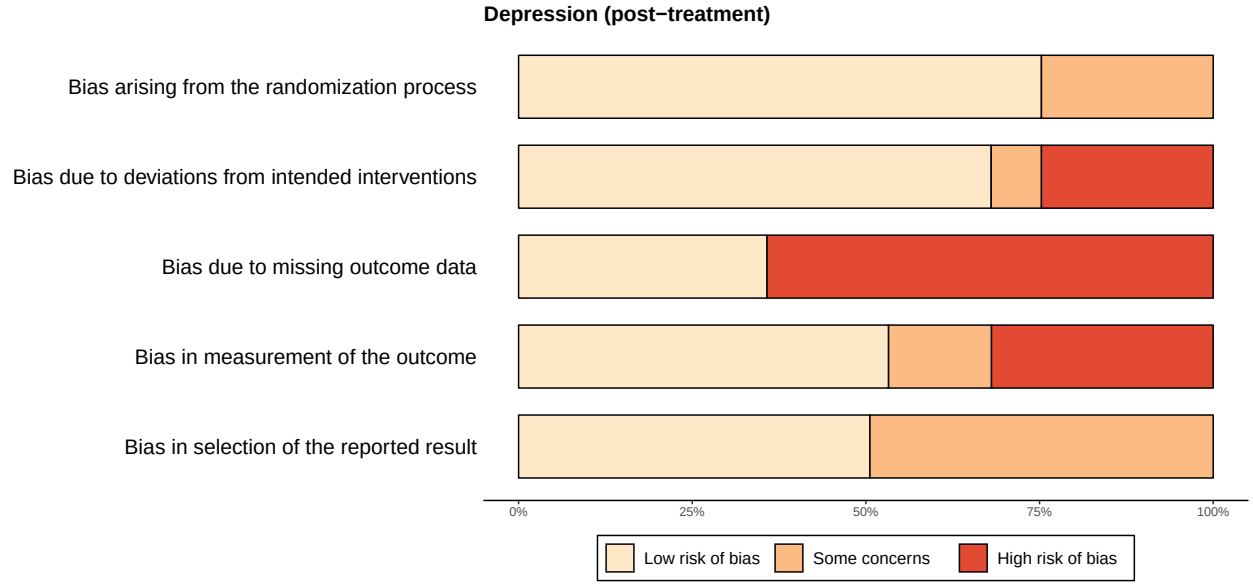

Figure F8. Risk of bias inherent in the summary effect for depression (post-treatment). Study-level biases are weighted according to the meta-analytic weights.

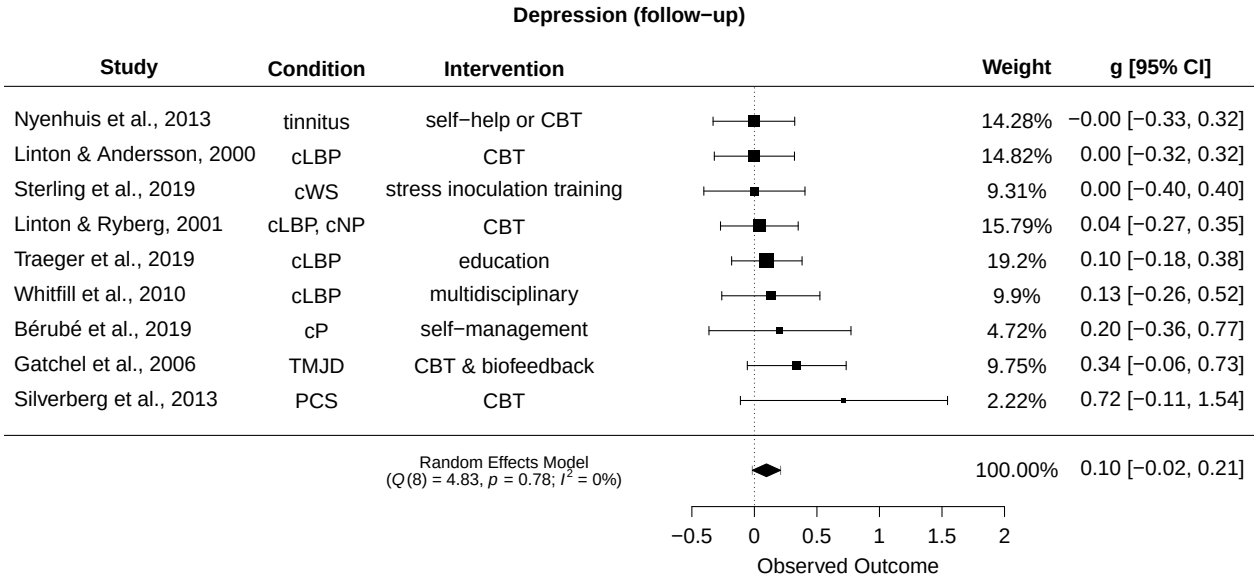

Figure F9. Forest plot of depression (follow-up). $g > 0$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. cP: Chronic pain. cWS: Chronic whiplash syndrome. PCS: Post-concussion syndrome. TMJD: Temporomandibular joint disorder

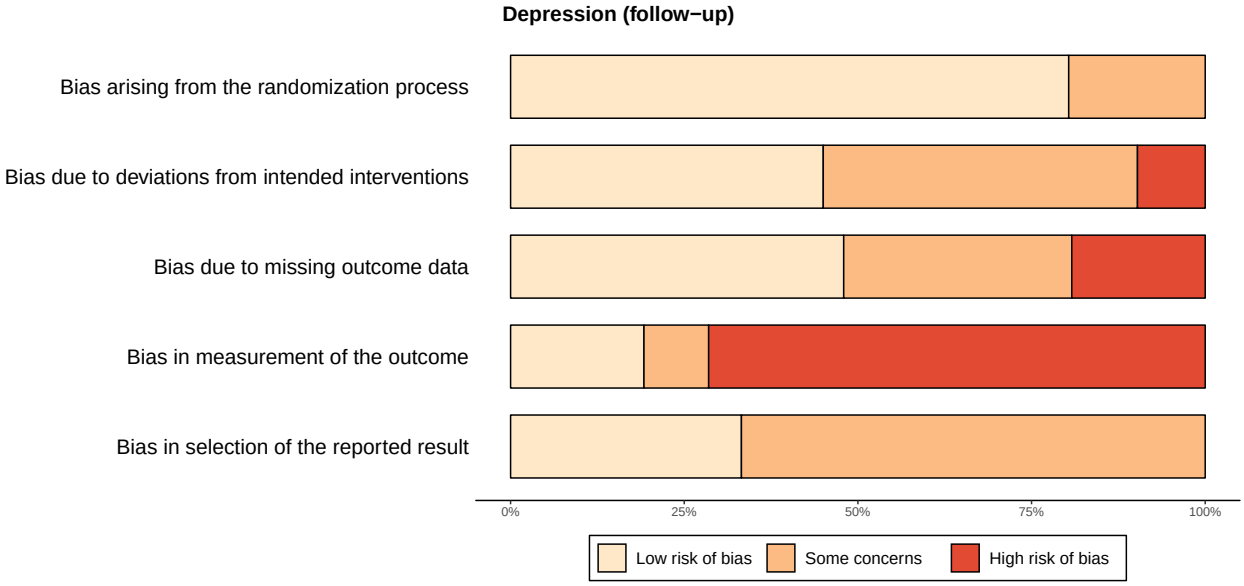

Figure F10. Risk of bias inherent in the summary effect for depression (follow-up). Study-level biases are weighted according to the meta-analytic weights.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = -0.27$ (95%-CI: [-0.6, 0.064]). The 3PSM revealed a corrected effect estimate of $g = 0.14$. A confidence interval could not be computed due to problems with model convergence. A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 1.65$, $p = .2$).

Health care utilization. Outcome data for studies measuring health care utilization are listed in Table E6.

Post-treatment. Out of four studies measuring health care utilization post-treatment, effect size data were available for none of them.

Follow-up. Out of eight studies measuring health care utilization at follow-up, effect size data were available for three studies ($n = 283$). Follow-up length ranged from 1.5 months to 12 months (*Median* = 10.5). There was a positive small and significant effect ($g = 0.31$, 95%-CI: [0.18, 0.44], see Figure F11). Heterogeneity was not significantly different from zero ($Q(2) = 0.13$, $p = .94$) and inconsistency was small to substantial ($I^2 = 0\%$, 95%-CI: [0%, 76.4%]). The resulting 95%-prediction interval ranged from 0.18 to 0.44.

Risk of bias in individual studies. Figure F12 summarizes the risk of bias ratings. See Figure E6 for risk of bias ratings for each study.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.26$ (95%-CI: [0.15, 0.38]). The 3PSM revealed a corrected effect estimate of $g = 0.82$. A confidence interval could not be computed due to problems with model convergence. A likelihood-ratio test revealed a significantly better fit of the 3PSM to the data ($\chi^2(1) = 5.75$, $p = .016$).

Consumer satisfaction. Outcome data for studies measuring consumer satisfaction are listed in Table E7.

Post-treatment. Out of six studies measuring consumer satisfaction post-treatment, appropriate effect size data were available for two studies ($n = 371$). Therefore, the data were synthesized narratively. In the study by Damush et al. (2003b), subjects with acute low back pain participated in a self-management program while control subjects received SC/TAU. Based on a sample of $n = 163$, there was no significant effect of the intervention ($g = -0.02$, 95%-CI: [-0.33, 0.29]).

In the study by Nyenhuis, Zastrutzki, Weise, Jäger, and Kröner-Herwig (2013), subjects

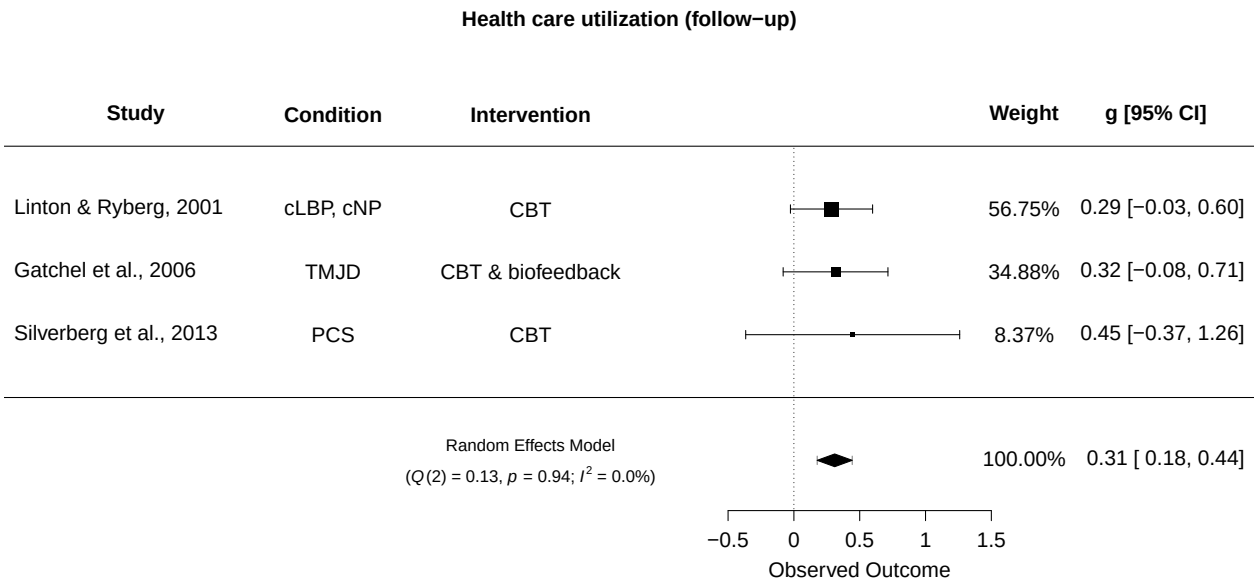

Figure F11. Forest plot of health care utilization (follow-up). $g > 1$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. PCS: Post-concussion syndrome. TMJD: Temporomandibular joint disorder.

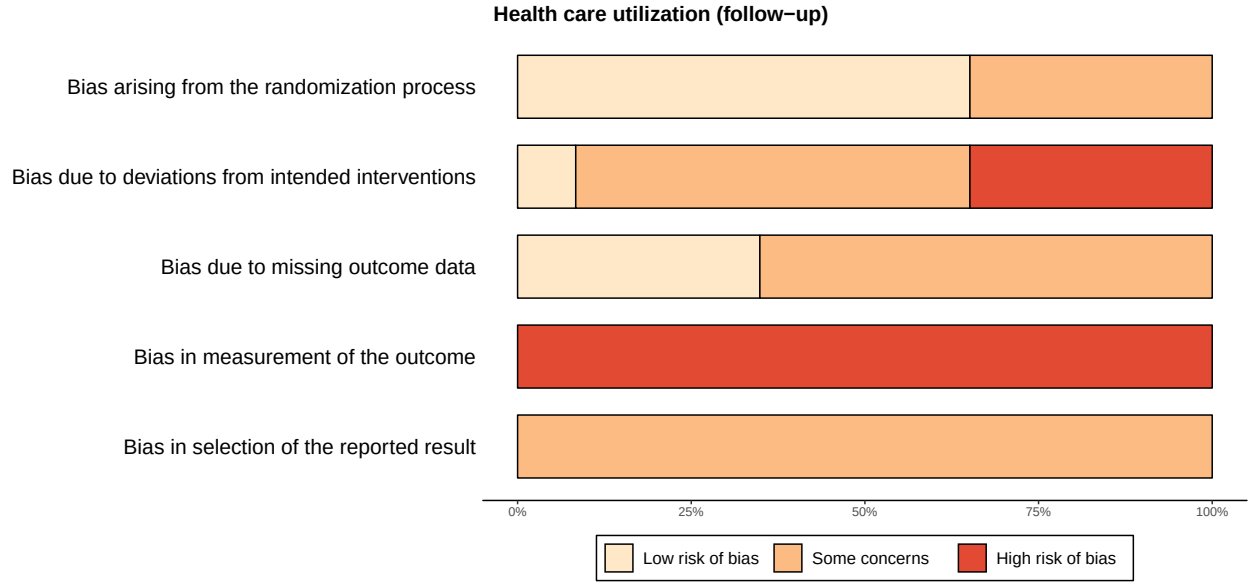

Figure F12. Risk of bias inherent in the summary effect for health care utilization (follow-up). Study-level biases are weighted according to the meta-analytic weights.

suffering from acute tinnitus were treated either with group CBT, bibliotherapy, or an online self-help program in the intervention groups. Except for an information sheet concerning the auditory system, tinnitus and treatment options, the control subjects received no treatment. There was a large significant combined effect of the interventions ($g = 1.21$, 95%-CI: [0.89, 1.54], $n = 208$).

Although the other studies did not provide appropriate data for meta-analytic integration, there was other information concerning the consumer satisfaction available. In the study by Gil-Jardiné et al. (2018) evaluating EMDR or reassurance compared to SC/TAU in patients at high risk for post-concussion syndrome, consumer satisfaction was rated on an 11-point numeric rating scale ranging from 0 to 10 with higher values indicating higher satisfaction. There was a median satisfaction of 9.5 (interquartile range (*IQR*): 8 - 10, $n = 34$) in the EMDR group, a median satisfaction of 8.5 (*IQR*: 7.25 - 10, $n = 38$) in the reassurance group and a median satisfaction of 8 (*IQR*: 6 - 10, $n = 37$) in the control group.

In the study by Karjalainen et al. (2004), consumer satisfaction was rated on the same scale. Subjects were patients suffering from subacute low back pain. The intervention consisted of advice, physiotherapeutic exercises and for a subset of subjects also of a worksite visit by a physiotherapist and a physician. The control group received SC/TAU. Intervention groups resulted in a combined mean satisfaction of 6.15 (range: 0 - 10, $n = 104$), while the SC/TAU group resulted in a mean satisfaction of 4.1 (range: 0 - 10, $n = 56$).

Follow-up. Out of five studies measuring consumer satisfaction at follow-up, effect size data were available for one study. In the study by Damush et al. (2003b, described above), there was no significant difference between the intervention and the control group ($g = 0.098$, 95%-CI: [-0.24, 0.43], $n = 139$) after a follow-up length of 11.25 months.

There were two further studies with relevant data, although they did not report enough data for calculating an effect size. In the study by Karjalainen et al. (2004, described above) there was a combined mean satisfaction of 5.99 (range: 0 - 10, $n = 103$) in the intervention groups and a mean satisfaction of 4.3 (range: 0 - 10, $n = 53$) in the SC/TAU group after at the 24-months follow-up.

In the study by Silverberg et al. (2013), subjects at risk for post-concussion syndrome

received six sessions of CBT. Consumer satisfaction was assessed using a 5-point Likert scale. At 1.5 months follow-up, the mean satisfaction in the intervention group was 4.69 (SD = 0.48, n = 13) indicating high satisfaction. There were no data available for the SC/TAU control group.

Online Supplement G

Additional analyses

Primary outcomes

Somatic symptom severity.

Post-treatment. Intervention intensity did not significantly moderate the treatment effect ($F(1,11) = 0.16$, $p = .7$, $R^2 = 0\%$). No moderator analysis of mean symptom duration could be computed as there were too few observations ($k = 2$). Type of population significantly moderated the treatment effect ($F(1,11) = 7.14$, $p = .022$, $R^2 = 63.1\%$). Specifically, there was no significant effect of studies with prevention populations (g = -0.16, 95%-CI: [-0.42, 0.11]), while there was a significant effect for studies with early intervention populations ($g = 0.23$, 95%-CI: [0.048, 0.42]). Type of control group did not significantly moderate the treatment effect ($F(3,9) = 2.82$, $p = .1$, $R^2 = 44.7\%$).

Descriptive analyses revealed a medium-sized interdependence between intervention intensity and type of population ($V = .35$) resulting from high intensity interventions being over-represented in early intervention populations. There was a large interdependence between intervention intensity and type of control group ($V = .56$) with all no treatment controls being compared to low intensity intensity interventions and all wait-list controls being compared to high intensity interventions. There was a medium-sized interdependence between type of population and type of control group ($V = .48$) with all no treatment and wait-list comparisons being conducted in early intervention populations.

Follow-up. Intervention intensity did not significantly moderate the treatment effect ($F(1,15) = 0.035$, $p = .85$, $R^2 = 0\%$). Mean symptom duration did not significantly moderate the treatment effect ($F(1,1) = 1.04$, $p = .49$, $R^2 = 0\%$). Type of population did not significantly moderate the treatment effect ($F(1,15) = 1.59$, $p = .23$, $R^2 = 25\%$). Type of control group did not significantly moderate the treatment effect ($F(2,14) = 0.55$, $p = .59$, $R^2 = 0\%$). Length of follow-up did not significantly moderate the treatment effect ($F(1,15) = 0.45$, $p = .5$, $R^2 = 0\%$).

There was a medium-sized negative correlation between intervention intensity and mean symptom duration (r_b = -.49) with high intensity interventions displaying lower mean symptom durations. There was a small interdependence between intervention intensity and

type of population (V = .17) resulting from high intensity interventions being under-represented in prevention populations and over-represented in early intervention populations. There was a medium-sized interdependence between intervention intensity and type of control group (V = .4) with high intensity interventions being over-represented in studies with SC/TAU controls. There was a medium-sized correlation between intervention intensity and length of follow-up (r_b = .35) with high intensity interventions displaying bigger lengths of follow-up. No rank correlation between mean symptom duration and type of population could be computed, as all studies providing mean symptom duration data were conducted in early intervention populations. There was a large negative rank correlation between mean symptom duration and type of control group (ρ = -.87). There was a medium-sized negative correlation between mean symptom duration and length of follow-up (r = -.49). There was a medium-sized interdependence between type of population and type of control group (V = .45) with all no treatment comparisons being conducted in early intervention populations. There was a medium-sized positive rank correlation between type of population and length of follow-up (ρ = .45). There was a small positive rank correlation between type of comparison and length of follow-up (ρ = .12).

Health-related quality of life.

Post-treatment. Intervention intensity did not significantly moderate the treatment effect ($F(1,9) = 0.061$, p = .81, R^2 = 0%). No moderator analysis of mean symptom duration could be conducted as there were too few observations (k = 1). Type of population significantly moderated the treatment effect ($F(1,9) = 6.91$, p = .027, R^2 = 80%). Specifically, there was no significant effect in prevention populations (g = -0.18, 95%-CI: [-0.51, 0.14], while there was a significant effect in early intervention populations (g = 0.24, 95%-CI: [0.073, 0.4]. Type of control group did not significantly moderate the treatment effect ($F(3,7)$ = 0.95, p = .47, R^2 = 0%).

There was a medium-sized interdependence between intervention intensity and type of population (V = .39) resulting from high intensity interventions being over-represented in early intervention populations. There was a large interdependence between intervention intensity and type of control group (V = .59) with all no treatment controls being compared to

low intensity interventions and all wait-list and placebo controls being compared to high intensity interventions. There was a large interdependence between type of population and type of control group (V = .57) with prevention populations being investigated in studies with SC/TAU controls, only.

Follow-up. Intervention intensity did not significantly moderate the treatment effect ($F(1,10) = 3.68$, $p = .08$, $R^2 = 100\%$). No moderator analysis of mean symptom duration could be conducted as there were too few observations ($k = 1$). Type of population significantly moderated the treatment effect ($F(1,10) = 6.14$, $p = .033$, $R^2 = 100\%$). Specifically, there was no significant effect in prevention populations (g = -0.028, 95%-CI: [-0.21, 0.15]), while there was a significant effect in early intervention populations (g = 0.21, 95%-CI: [0.091, 0.34]). Type of control group did not significantly moderate the treatment effect ($F(2,9) = 2.56$, $p = .13$, $R^2 = 100\%$). Length of follow-up did not significantly moderate the treatment effect ($F(1,10) = 1.26$, $p = .29$, $R^2 = 61.3\%$).

There was a small interdependence between intervention intensity and type of population (V = .29) resulting from high intensity interventions being over-represented in early intervention populations. There was a large interdependence between intervention intensity and type of control group (V = .56) with all no treatment controls being compared to low intensity interventions. There was a large correlation between intervention intensity and length of follow-up (r_b = .63). There was a medium-sized interdependence between type of population and type of control group (V = .36) with all no treatment comparisons being conducted in early intervention populations. There was a large positive rank correlation between type of population and length of follow-up (ρ = .52). There was a small positive rank correlation between type of comparison and length of follow-up (ρ = .19).

Secondary outcomes

Unwanted negative treatment effects. No additional analyses of unwanted negative treatment effects could be conducted, since these data were synthesized narratively.

Diagnostic status concerning SSD/FSS.

Post-treatment. Intervention intensity did not significantly moderate the treatment effect ($F(1,2) = 1.9$, $p = .3$, $R^2 = 38.4\%$). No moderator analysis of mean symptom duration

could be computed as there were no observations. Type of population did not significantly moderate the treatment effect ($F(1,2) = 1.9$, $p = .3$, $R^2 = 38.4\%$). Type of control group did not significantly moderate the treatment effect ($F(2,1) = 0.68$, $p = .65$, $R^2 = 0\%$).

Descriptive analyses revealed a perfect interdependence between intervention intensity and type of population ($V = 1$) with all studies with prevention populations evaluating low intensity interventions and all studies with early intervention populations evaluating high intensity interventions. There was a perfect interdependence between intervention intensity and type of control group ($V = 1$) with high intensity interventions being only evaluated in studies with wait-list and placebo controls and low intensity interventions being evaluated in studies with SC/TAU controls, only. There was a perfect interdependence between type of population and type of control group ($V = 1$) with all studies with early intervention populations using wait-list and placebo controls, while all studies with prevention populations were conducted with SC/TAU controls.

Follow-up. No additional analyses could be conducted, since there were too few studies with available data ($k = 2$).

Anxiety.

Post-treatment. No moderator analysis of intervention intensity could be conducted as all studies examined high intensity interventions. No moderator analysis of mean symptom duration could be conducted as no study provided data for this moderator. Type of population did not significantly moderate the treatment effect ($F(1,1) = 68.1$, $p = .077$, $R^2 = 0\%$). No moderator analysis of type of control group could be conducted as all studies examined SC/TAU controls.

Follow-up. No moderator analysis of intervention intensity could be conducted as all studies examined high intensity interventions. No moderator analysis of mean symptom duration could be conducted as no study provided data for this moderator. Type of population did not significantly moderate the treatment effect ($F(1,4) = 0.005$, $p = .95$, $R^2 = 0\%$). No moderator analysis of type of control group could be conducted as all studies examined SC/TAU controls. Length of follow-up did not significantly moderate the treatment effect ($F(1,4) = 0.3$, $p = .62$, $R^2 = 0\%$).

There was a large positive rank correlation between type of population and length of follow-up (ρ = .84).

Depression.

Post-treatment. Intervention intensity did not significantly moderate the treatment effect ($F(1,3) = 1.34$, $p = .33$, $R^2 = 0\%$). No moderator analysis of mean symptom duration could be computed as there were too few observations ($k = 2$). Type of population did not significantly moderate the treatment effect ($F(1,3) = 0.22$, $p = .67$, $R^2 = 0\%$). Type of control group did not significantly moderate the treatment effect ($F(2,2) = 0.6$, $p = .62$, $R^2 = 0\%$).

There was a small interdependence between intervention intensity and type of population ($V = .25$) with all studies with prevention populations evaluating high intensity interventions. There was a perfect interdependence between intervention intensity and type of control group ($V = 1$) with high intensity interventions being only evaluated in studies with SC/TAU and placebo controls and low intensity interventions being evaluated in studies with no treatment controls, only. There was a large interdependence between type of population and type of control group ($V = .61$) with prevention populations being investigated in studies with SC/TAU controls, only.

Follow-up. Intervention intensity did not significantly moderate the treatment effect ($F(1,7) = 0.67$, $p = .44$, $R^2 = 99.4\%$). Mean symptom duration did not significantly moderate the treatment effect ($F(1,1) = 1.71$, $p = .19$, $R^2 = 0\%$). Type of population did not significantly moderate the treatment effect ($F(1,7) = 2.83$, $p = .14$, $R^2 = 100\%$). Type of control group did not significantly moderate the treatment effect ($F(2,6) = 0.086$, $p = .92$, $R^2 = 52.4\%$). Length of follow-up did not significantly moderate the treatment effect ($F(1,7) = 0.84$, $p = .39$, $R^2 = 100\%$).

There was a medium-sized negative correlation between intervention intensity and mean symptom duration (r_b = -.49). There was a small interdependence between intervention intensity and type of population ($V = .19$). There was a large interdependence between intervention intensity and type of control group ($V = .66$) with all SC/TAU and placebo comparisons being conducted in studies evaluating high intensity interventions. There was a small negative correlation between intervention intensity and length of follow-up (r_b = -.11).

No rank correlation between mean symptom duration and type of population could be computed, since all studies providing mean symptom duration data were conducted in early intervention populations. There was a large negative rank correlation between symptom duration and type of comparison (ρ = -.87). There was a nearly perfect positive correlation between mean symptom duration and length of follow-up (r = .99). There was a medium-sized interdependence between type of population and type of control group (V = .38) with all no treatment and placebo comparisons being conducted in studies with early intervention populations. There was a large positive rank correlation between type of population and length of follow-up (ρ = .58). There was a small negative rank correlation between type of comparison and length of follow-up (ρ = -.29).

Health care utilization.

Post-treatment. No additional analyses could be conducted, since there were no studies with available data.

Follow-up. No moderator analysis of intervention intensity could be computed as all studies evaluated high intensity interventions. No moderator analysis of mean symptom duration could be computed as there were too few available studies (k = 1). Type of population did not significantly moderate the treatment effect ($F(1,1) = 8.46$, p = .21, R^2 = 0%). Type of control group did not significantly moderate the treatment effect ($F(1,1) = 0.012$, p = .93, R^2 = 0%). Length of follow-up did not significantly moderate the treatment effect ($F(1,1) = 145.3$, p = .053, R^2 = 0%).

There was a large interdependence between type of population and type of control group (V = .5) with all no treatment comparisons being conducted in studies with early intervention populations. There was a large positive rank correlation between type of population and length of follow-up (ρ = .87). There was a no rank correlation between type of comparison and length of follow-up (ρ = 0).

Consumer satisfaction. No additional analyses could be conducted, since these data have been synthesized narratively.

Online Supplement H

Sensitivity analyses: exclusion of cluster-randomized trials

Primary outcomes

Somatic symptom severity.

Post-treatment. Out of 18 studies measuring somatic symptom severity post-treatment, effect size data were available for 12 studies (n = 1,944). There was a small and non-significant effect ($g = 0.12$, 95%-CI: [-0.079, 0.31], see Figure H1). Heterogeneity was significantly different from zero ($Q(11) = 37.8$, $p < .0001$) and inconsistency was moderate to considerable ($I^2 = 70\%$, 95%-CI: [37.6%, 90.3%]). The resulting 95%-prediction interval ranged from -0.46 to 0.7.

Risk of bias in individual studies. Risk of bias inherent in the summary effect is depicted in Figure H2.

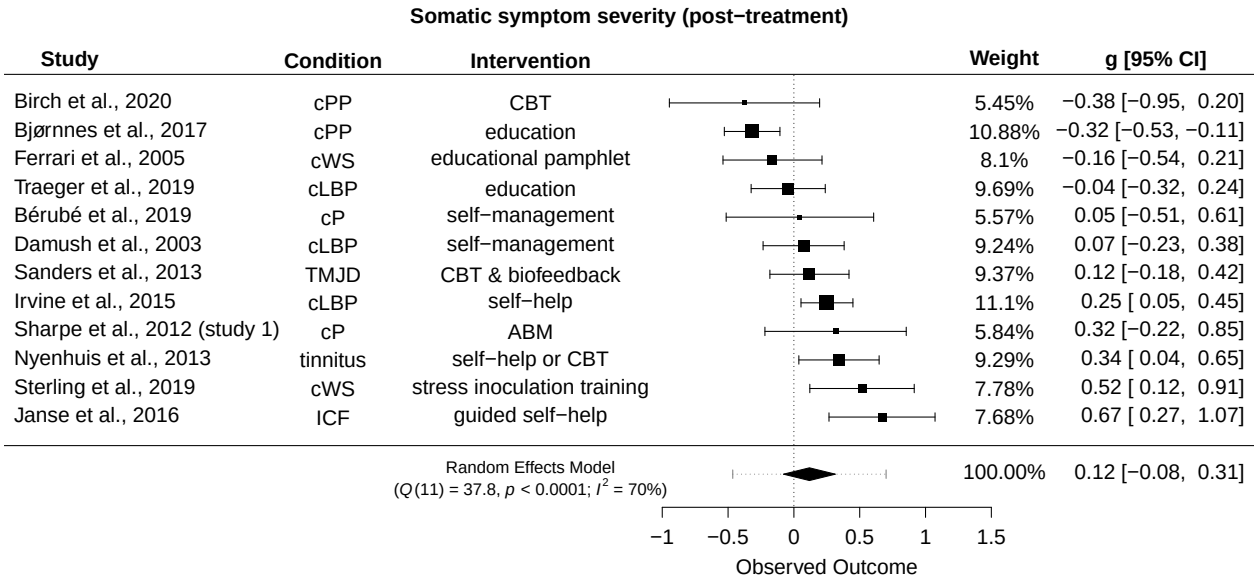

Figure H1. Forest plot of somatic symptom severity (post-treatment). $g > 0$ indicates more favorable outcomes in the intervention group. ABM: Attention bias modification. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cP: Chronic pain. cPP: Chronic postoperative pain. cWS: Chronic whiplash syndrome. ICF: Idiopathic chronic fatigue. TMJD: Temporomandibular joint disorder.

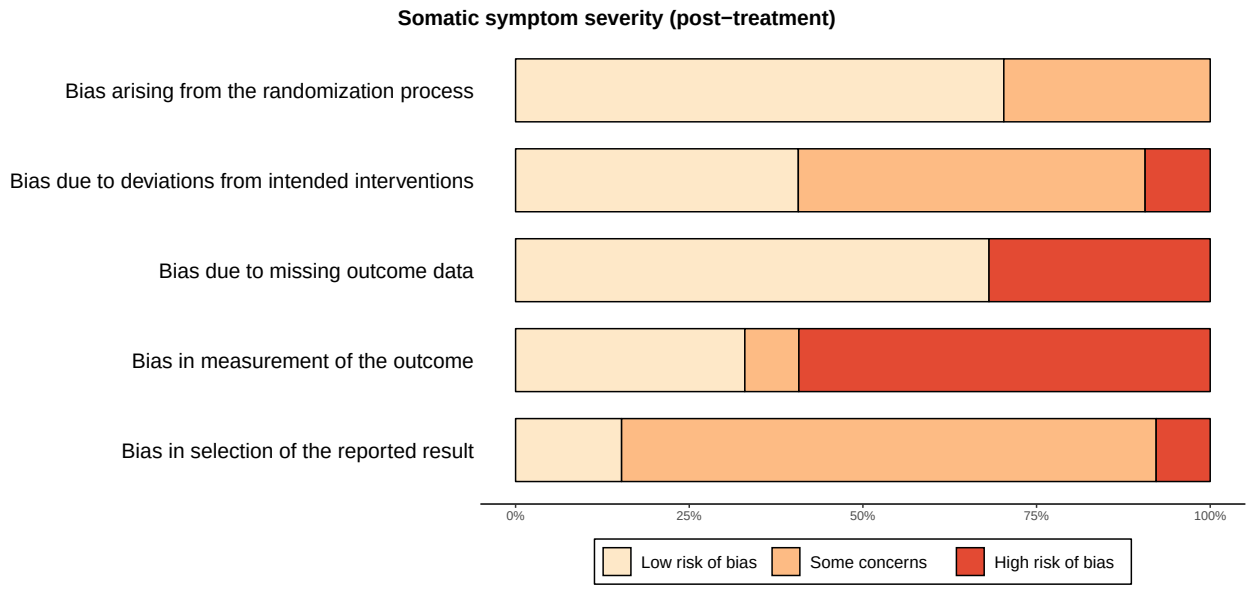

Figure H2. Risk of bias inherent in the summary effect for somatic symptom severity (post-treatment). Study-level biases are weighted according to the meta-analytic weights.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = -0.073$ (95%-CI: [-0.68, 0.53]). The 3PSM revealed a corrected effect estimate of $g = -0.023$ (95%-CI: [-0.2, 0.16]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 2.64$, $p = .1$).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,10) = 0.24$, $p = .64$, $R^2 = 0\%$). No moderator analysis of mean duration of symptoms could be conducted as there were too few observations ($k = 2$). Type of population significantly moderated the treatment effect ($F(1,10) = 7.4$, $p = .022$, $R^2 = 62.9\%$). Specifically, the effects in studies with prevention populations were not significantly deviating from zero ($g = -0.16$, 95%-CI: [-0.43, 0.11]), while there was a significant effect for studies with early intervention populations ($g = 0.25$, 95%-CI: [0.055, 0.44]). Type of control group did not significantly moderate the treatment effect ($F(3,8) = 2.43$, $p = .14$, $R^2 = 42.5\%$).

Descriptive analyses revealed a medium-sized interdependence between intervention intensity and type of population ($V = .31$) resulting from high intensity interventions being over-represented in studies with early intervention populations. There was a large interdependence between intervention intensity and type of control group ($V = .56$) with all

no treatment comparisons being conducted in studies with low intensity interventions and all wait-list control comparisons being conducted in studies with high intensity interventions. There was a large interdependence between type of population and type of control group ($V = .56$) with all no treatment and wait-list comparisons being conducted in studies with early intervention populations.

Follow-up. Out of 23 studies measuring somatic symptom severity at follow-up, effect size data were available for 16 studies (n = 2,346). Follow-up length ranged from 1.5 months to 12 months (*Median* = 9.25). There was a small and significant positive effect ($g = 0.25$, 95%-CI: [0.088, 0.41], see Figure H3). Heterogeneity was significantly different from zero ($Q(15)$ = 37.4, $p = .001$) and inconsistency was small to considerable ($I^2 = 60.8\%$, 95%-CI: [27.9%, 88%]). The resulting 95%-prediction interval ranged from -0.23 to 0.73.

Risk of bias in individual studies. Risk of bias inherent in the summary effect is depicted in Figure H4.

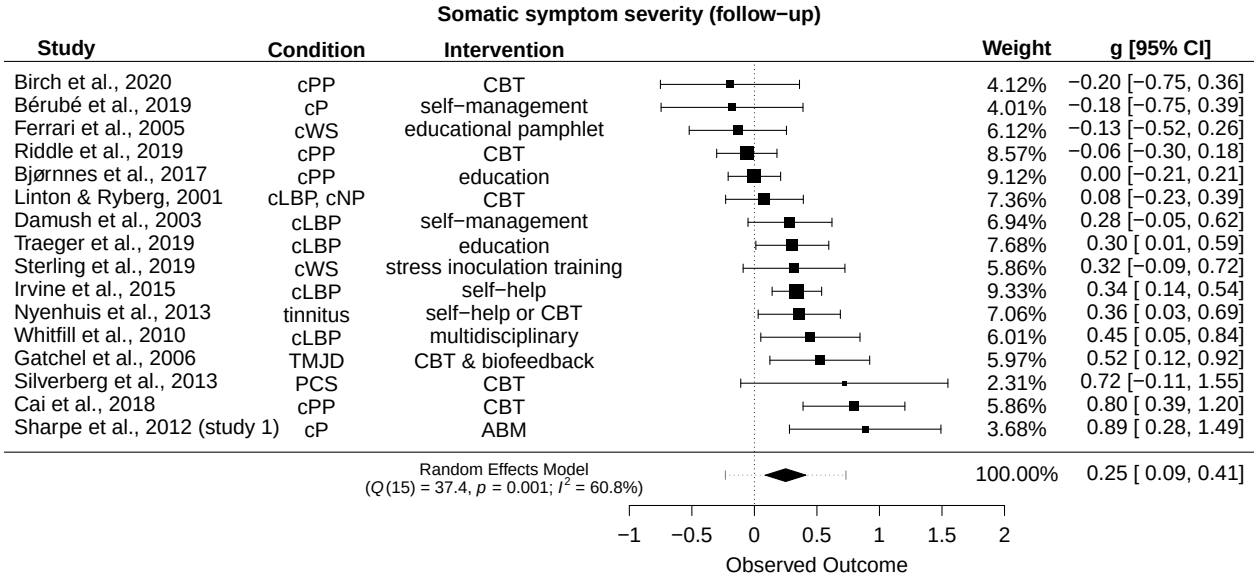

Figure H3. Forest plot of somatic symptom severity (follow-up). $g > 0$ indicates more favorable outcomes in the intervention group. ABM: Attention bias modification. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. cP: Chronic pain. cPP: Chronic postoperative pain. cWS: Chronic whiplash syndrome. PCS: Post-concussion syndrome. TMJD: Temporomandibular joint disorder.

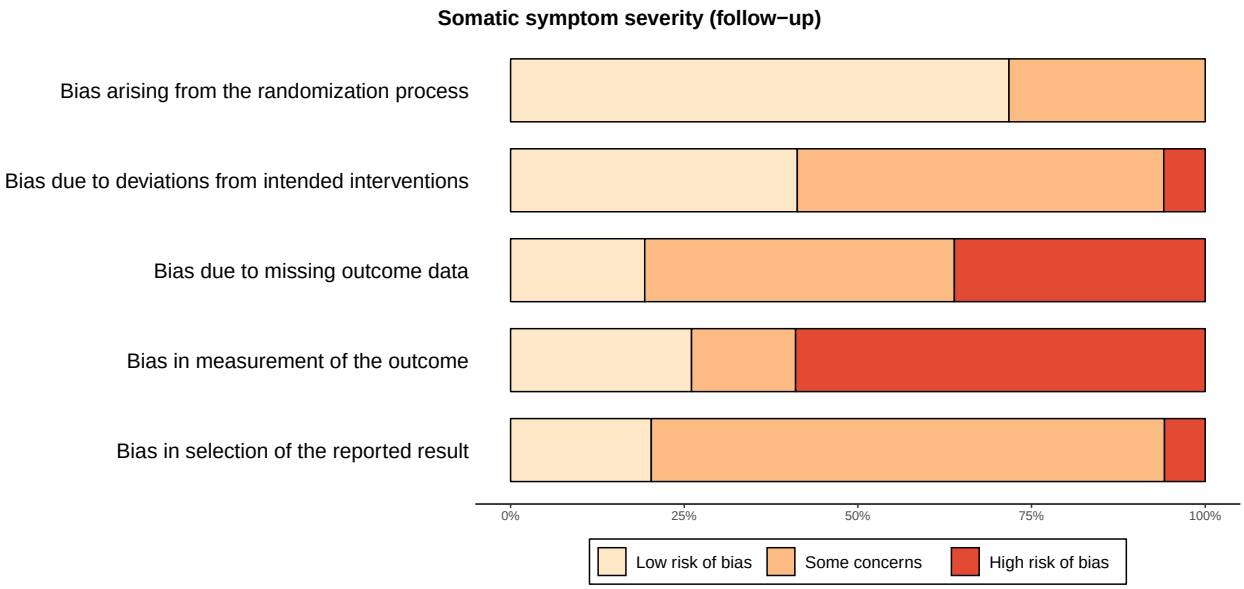

Figure H4. Risk of bias inherent in the summary effect for somatic symptom severity (follow-up). Study-level biases are weighted according to the meta-analytic weights.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.013$ (95%-CI: [-0.39, 0.42]). The 3PSM revealed a corrected effect estimate of $g = 0.12$ (95%-CI: [-0.054, 0.3]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 2.36$, $p = .12$).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,14) = 0.028$, $p = .87$, $R^2 = 0\%$). Mean symptom duration did not significantly moderate the treatment effect ($F(1,1) = 1.04$, $p = .49$, $R^2 = 0\%$). Type of population did not significantly moderate the treatment effect ($F(1,14) = 1.43$, $p = .25$, $R^2 = 23.2\%$). Type of control group did not significantly moderate the treatment effect ($F(2,13) = 0.52$, $p = .61$, $R^2 = 0\%$). Length of follow-up did not significantly moderate the treatment effect ($F(1,14) = 0.67$, $p = .43$, $R^2 = 0\%$).

There was a medium-sized negative correlation between intervention intensity and mean symptom duration ($r_b = -.49$) with high intensity interventions displaying lower mean symptom durations. There was a small interdependence between intervention intensity and type of population ($V = .13$) resulting from high intensity interventions being slightly over-represented in studies with early intervention populations. There was a medium-sized

interdependence between intervention intensity and type of control group ($V = .38$) with high intensity interventions being over-represented in studies with SC/TAU controls. There was a medium-sized correlation between intervention intensity and length of follow-up ($r_b = .34$) with high intensity interventions displaying bigger lengths of follow-up. No rank correlation between mean symptom duration and type of population could be computed as all studies providing mean symptom duration data were conducted in early intervention populations. There was a large negative rank correlation between mean symptom duration and type of control group ($\rho = -.87$). There was a medium-sized negative correlation between mean symptom duration and length of follow-up ($r = -.49$). There was a medium-sized interdependence between type of population and type of control group ($V = .48$) with all no treatment comparisons being conducted in studies with early intervention populations. There was a medium-sized positive rank correlation between type of population and length of follow-up ($\rho = .4$). There was a small positive rank correlation between type of comparison and length of follow-up ($\rho = 0.13$).

Health-related quality of life.

Post-treatment. Out of 15 studies measuring health-related quality of life post-treatment, effect size data were available for nine studies (n = 1,333). There was a small and non-significant effect ($g = 0.13$, 95%-CI: [-0.11, 0.37], see Figure H5). Heterogeneity was significantly different from zero ($Q(8) = 19.6$, $p = .012$) and inconsistency was small to considerable ($I^2 = 61.2\%$, 95%-CI: [14.2%, 93.6%]). The resulting 95%-prediction interval ranged from -0.42 to 0.68.

Risk of bias in individual studies. Risk of bias inherent in the summary effect is depicted in Figure H6.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of g = 0.33 (95%-CI: [-0.007, 0.66]). The 3PSM revealed a corrected effect estimate of $g = 0.035$ (95%-CI: [-0.16, 0.23]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 2.26$, p = .13).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,7) = 0.011$, p = .92, R^2 = 0%). No moderator analysis of mean symptom

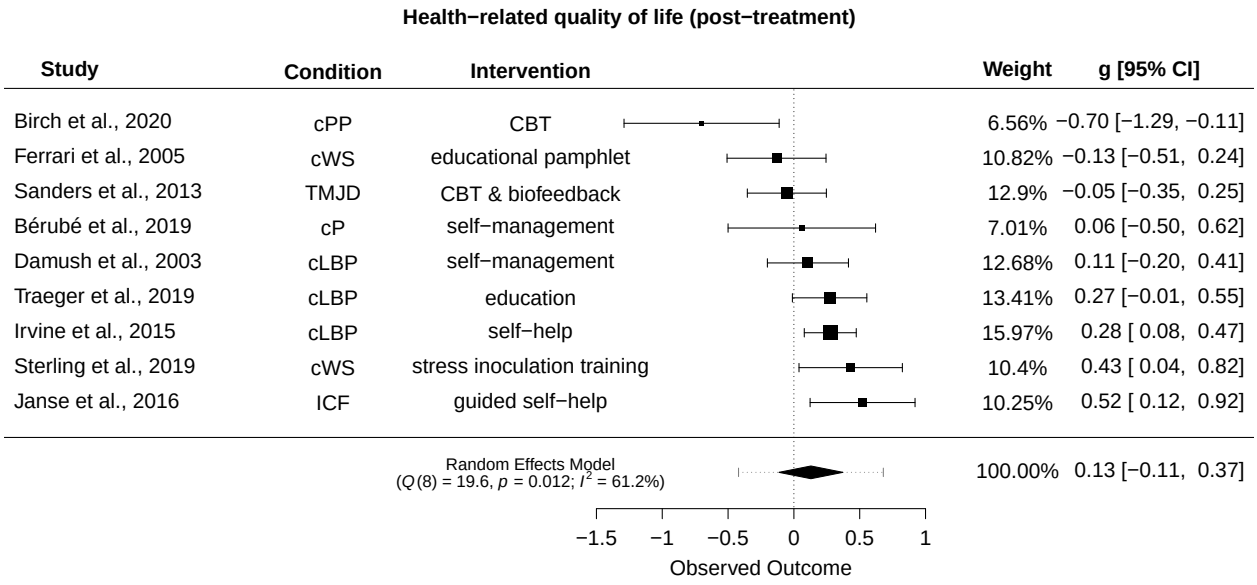

Figure H5. Forest plot of health-related quality of life (post-treatment). $g > 0$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cP: Chronic pain. cPP: Chronic postoperative pain. cWS: Chronic whiplash syndrome. ICF: Idiopathic chronic fatigue. TMJD: Temporomandibular joint disorder.

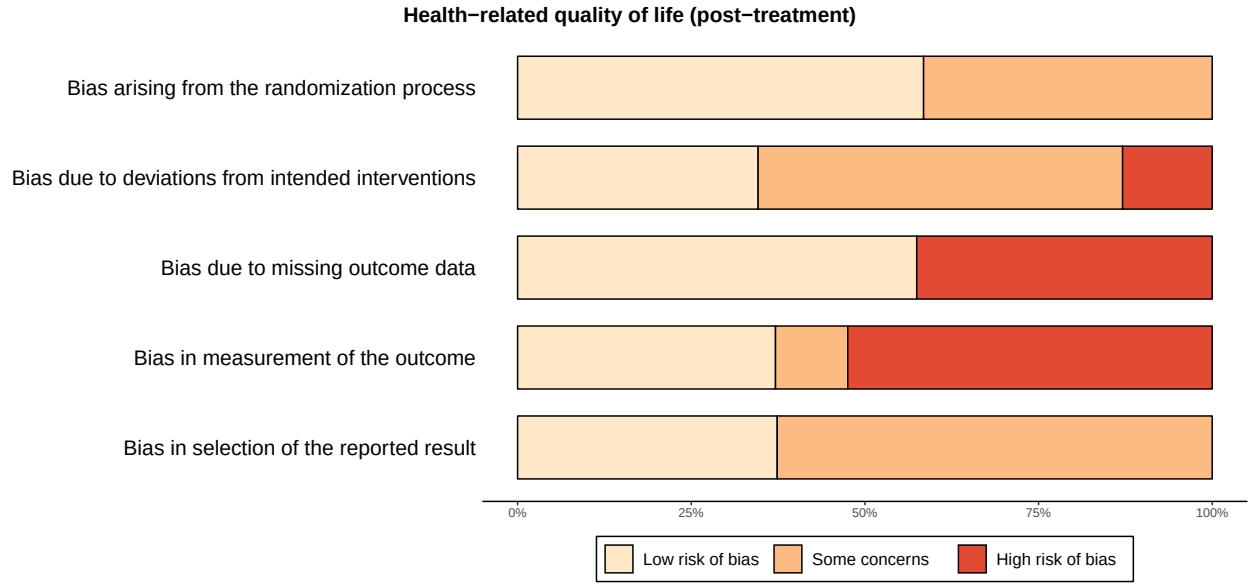

Figure H6. Risk of bias inherent in the summary effect for health-related quality of life (post-treatment). Study-level biases are weighted according to the meta-analytic weights.

duration could be conducted as there were too few observations ($k = 1$). Type of population significantly moderated the treatment effect ($F(1,7) = 5.64$, $p = .049$, $R^2 = 76.1\%$). Specifically, there was no significant effect in studies with prevention populations ($g = -0.22$, 95%-CI: [-0.62, 0.19]) while there was a significant effect in studies with early intervention populations ($g = 0.24$, 95%-CI: [0.039, 0.44]). Type of control group did not significantly moderate the treatment effect ($F(3,5) = 0.71$, $p = .59$, $R^2 = 0\%$).

There was a small interdependence between intervention intensity and type of population ($V = .19$) with high intensity interventions being over-represented in studies with early intervention populations. There was a large interdependence between intervention intensity and type of control group ($V = .7$) with all no treatment comparisons being conducted in studies with low intensity interventions and all wait-list and placebo comparisons being conducted in studies with high intensity interventions. There was a large interdependence between type of population and type of control group ($V = .63$) with prevention populations interventions being investigated in studies with SC/TAU controls, only.

Follow-up. Out of 17 studies measuring health-related quality of life at follow-up, effect size data were available for 11 studies ($n = 1{,}589$). Follow-up length ranged from 2 months to 12 months (*Median* = 9.5). There was a small non-significant effect ($g = 0.12$, 95%-CI: [-0.012, 0.25], see Figure H7). Heterogeneity was not significantly different from zero ($Q(10) = 13.1$, $p = .22$) and inconsistency was small to considerable ($I^2 = 25.6\%$, 95%-CI: [0%, 78.1%]). The resulting 95%-prediction interval ranged from -0.14 to 0.38.

Risk of bias in individual studies. Risk of bias is depicted in Figure H8.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.24$ (95%-CI: [0.028, 0.45]). The 3PSM revealed a corrected effect estimate of $g = 0.15$ (95%-CI: [-0.01, 0.31]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 0.36$, $p = .55$).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,9) = 3.48$, $p = .095$, $R^2 = 100\%$). No moderator analysis of mean symptom duration could be conducted as there were too few observations ($k = 1$). Type of population significantly moderated the treatment effect ($F(1,9) = 5.41$, $p = .045$, $R^2 = 96.8\%$).

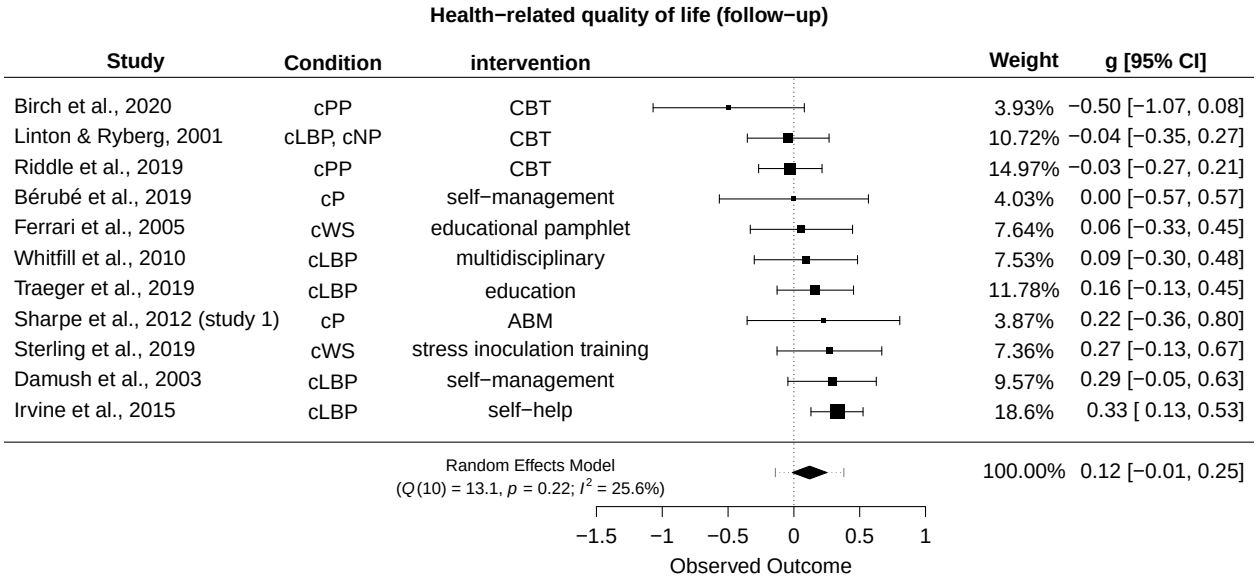

Figure H7. Forest plot of health-related quality of life (follow-up). $g > 0$ indicates more favorable outcomes in the intervention group. ABM: Attention bias modification. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. cP: Chronic pain. cPP: Chronic postoperative pain. WS: Chronic whiplash syndrome.

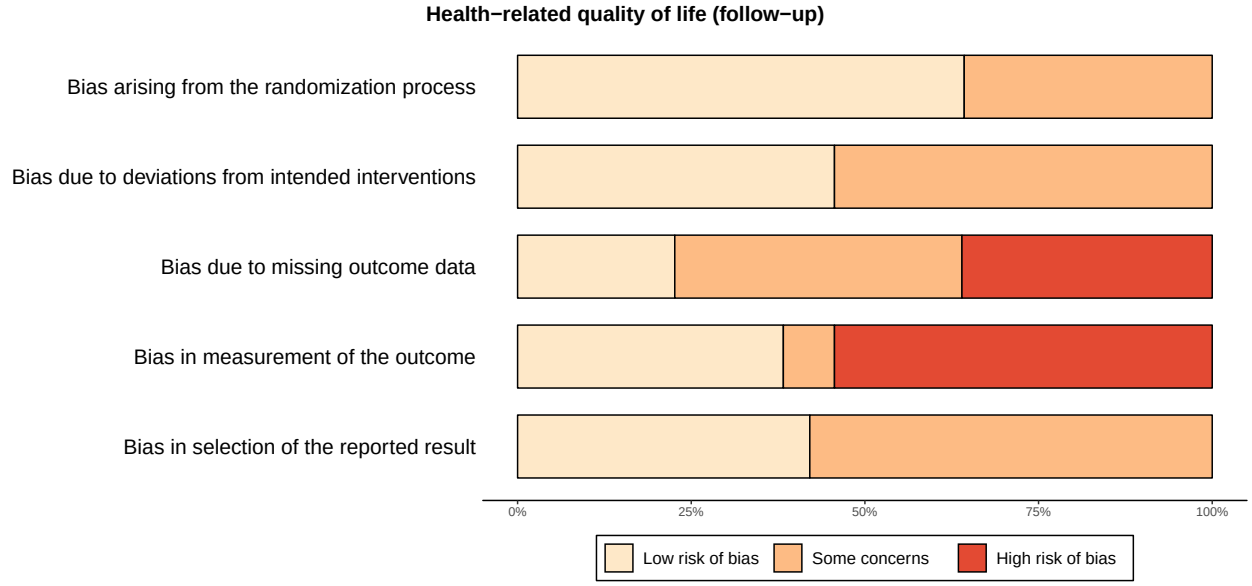

Figure H8. Risk of bias inherent in the summary effect for health-related quality of life (follow-up). Study-level biases are weighted according to the meta-analytic weights.

Specifically, there was no significant effect in studies with prevention populations (g = -0.028, 95%-CI: [-0.22, 0.16]) while there was a significant effect in studies with early intervention populations (g = 0.21, 95%-CI: [0.079, 0.34]). Type of control group did not significantly moderate the treatment effect ($F(2,8) = 2.37$, $p = .16$, $R^2 = 100\%$). Length of follow-up did not significantly moderate the treatment effect ($F(1,9) = 2.15$, $p = .18$, $R^2 = 83.6\%$).

There was a small interdependence between intervention intensity and type of population ($V = .26$) resulting from high intensity interventions being over-represented in studies with early intervention populations. There was a large interdependence between intervention intensity and type of control group ($V = .55$) with all no treatment comparisons being conducted in studies with low intensity interventions. There was a large correlation between intervention intensity and length of follow-up ($r_b = .81$). There was a medium-sized interdependence between type of population and type of control group ($V = .36$) with all no treatment comparisons being conducted in studies with early intervention populations. There was a medium-sized positive rank correlation between type of population and length of follow-up ($\rho = .47$). There was a small positive rank correlation between type of comparison and length of follow-up ($\rho = .27$).

Secondary outcomes

Unwanted negative treatment effects.

Post-treatment. Since only one study assessed unwanted negative treatment effects post-treatment, we describe these data narratively. Sterling, Smeets, Keijzers, Warren, Kenardy (2019) evaluated the effect of stress inoculation training in combination with guideline-based exercise compared to guideline-based exercise alone (SC/TAU) for patients suffering from whiplash-associated disorder (n = 108). The researchers assessed adverse effects (i.e., exacerbation of a pre-existing condition) and adverse events (i.e., events that are life-threatening, require inpatient hospitalization, or will result in persistent or significant disability or incapacity) via open ended questions. In each trial arm, one subject reported neck pain exacerbation, while no subject reported adverse events.

Follow-up. Only two studies assessed unwanted negative treatment effects at follow-up. Therefore, we describe these data narratively. In the study by Traeger et al. (2019), patients with acute low back pain (n = 202) were randomized to an intensive patient education condition or to a placebo education condition. Both treatments were delivered face-to-face. The researchers recorded adverse events during the trial. Over a follow-up time of 10.5 months, there were no reported adverse events in any of the treatment groups.

In the study by Riddle et al. (2019), patients scheduled for a knee arthroplasty at risk for chronic pain (n = 402) received either CBT-based pain coping skills training or arthritis education serving as placebo condition. Beyond that, there was a third trial arm providing SC/TAU, only. Unwanted negative treatment effects were assessed during data collection and by medical record review after a follow-up time of 10.5 months. There were no significant differences neither in adverse events (e.g., emergency room visits due to knee pain, psychological distress, elevated depressive symptoms) nor in serious adverse events (e.g., hospitalization, surgery, infection, death) between groups.

Diagnostic status concerning SSD/FSS.

Post-treatment. Four studies measured diagnostic status concerning SSD/FSS post-treatment. Effect size data were available for all of them (n = 427). A random-effects meta-analysis revealed a risk ratio of 0.92 (95%-CI: [0.62, 1.37], see Figure H9). Heterogeneity was not significantly different from zero ($Q(3) = 7.53$, $p = .057$) and inconsistency was small to considerable ($I^2 = 59.8\%$, 95%-CI: [0%, 97.5%]). The resulting 95%-prediction interval ranged from 0.45 to 1.89.

Risk of bias in individual studies. Figure H10 depicts the risk of bias inherent in the diagnostic status summary effect.

Meta-bias. The PET-PEESE revealed a corrected risk ratio of 0.68 (95%-CI: [0.12, 3.83]). The 3PSM revealed a corrected risk ratio of 0.96 (95%-CI: [0.66, 1.42]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 0.4$, p = .53).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,2) = 1.9$, p = .3, R^2 = 38.4%). No moderator analysis of mean symptom

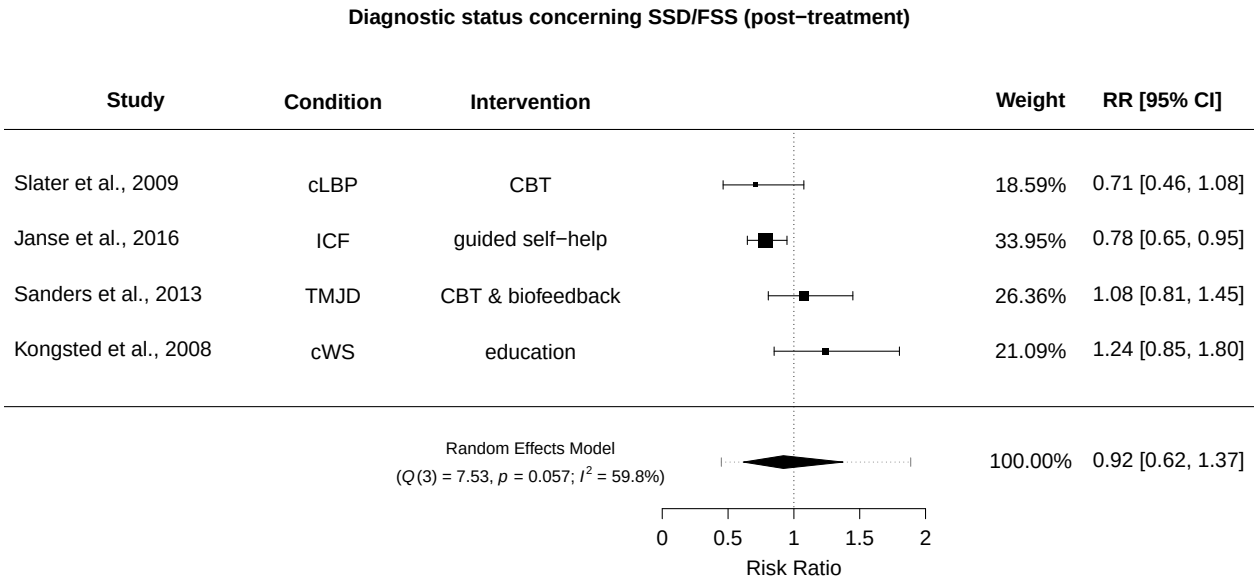

Figure H9. Forest plot of diagnostic status concerning SSD/FSS (post-treatment). *RR* < 1 indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cWS: Chronic whiplash syndrome. ICF: Idiopathic chronic fatigue. TMJD: Temporomandibular joint disorder.

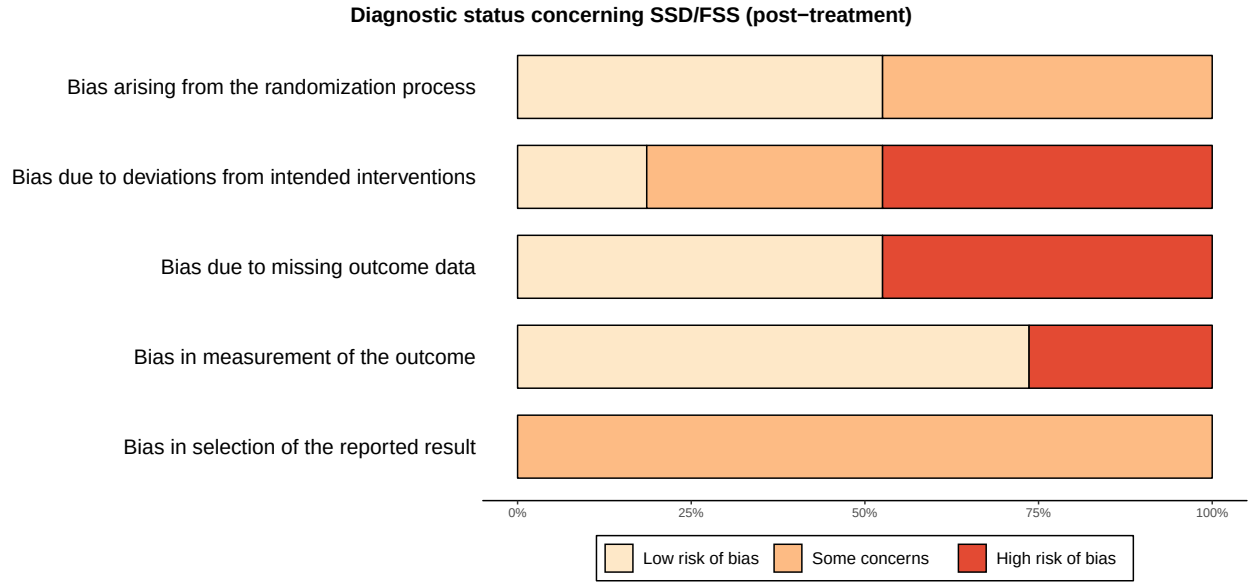

Figure H10. Risk of bias inherent in the summary effect for diagnostic status concerning SSD/FSS (post-treatment). Study-level biases are weighted according to the meta-analytic weights.

duration could be computed as there were no observations. Type of population did not significantly moderate the treatment effect ($F(1,2) = 1.9$, $p = .3$, $R^2 = 38.4\%$). Type of control group did not significantly moderate the treatment effect ($F(2,1) = 0.68$, $p = .65$, $R^2 = 0\%$). Descriptive analyses revealed a perfect interdependence between intervention intensity and type of population ($V = 1$) with all studies with prevention populations investigating low intensity interventions and all studies with early intervention populations evaluating high intensity interventions. There was a perfect interdependence between intervention intensity and type of control group ($V = 1$) with high intensity interventions being only evaluated in studies with wait-list and placebo controls and low intensity interventions being evaluated in studies with SC/TAU controls, only. There was a perfect interdependence between type of population and type of control group ($V = 1$) with all studies with early intervention populations using wait-list and placebo controls, while all studies with prevention populations were using SC/TAU controls.

Follow-up. Out of three studies measuring diagnostic status concerning SSD/FSS at follow-up, appropriate effect size data were available for two studies. Therefore, these data are synthesized narratively. In the study by Gil-Jardiné et al. (2018), patients at high risk for developing a postconcussion syndrome were treated with a session of either EMDR or reassurance by a therapist in the emergency room. Control subjects received SC/TAU. Diagnostic status was determined via an interview based on the DSM-IV criteria for postconcussion syndrome. Based on a sample of $n = 123$ and a follow-up length of 3 months, there was a significant effect favoring the intervention groups ($RR = 0.54$, 95%-CI: [0.37, 0.78]). It is important to note that this effect stems from a worst-case-scenario analysis in which subjects abandoning the intervention protocol due to early discharge or clinical worsening were considered as having an SSD/FSS at follow-up.

Kongsted et al. (2008) examined the effect of oral advice given by a nurse at a home visit to patients presenting with a whiplash injury compared to SC/TAU consisting of an educational pamphlet. These patients were of comparably lower risk for chronic whiplash syndrome since patients at high risk were invited to participate in another trial. Diagnosis was defined via a combination of a neck pain measure and current work status. Based on a sample

of 158 subjects and a follow-up length of 12 months, there was no significant effect of the intervention (*RR* = 1.2, 95%-CI: [0.93; 1.55]).

Although the study by Gatchel et al. (2006) did not provide appropriate effect size data for meta-analytic integration, it reports the effect of the intervention in another effect size metric. Therefore, we describe this study here, too. The study evaluated a combined CBT and biofeedback treatment program for patients suffering from acute jaw pain at high risk for developing a temporomandibular joint disorder. Patients in the control group received no intervention in the context of the trial. Diagnosis was determined by fulfilling the criteria for a pain disorder using the Structured Clinical Interview for DSM-IV. Based on a sample of n = 101 and a follow-up length of 10.5 months, there was a significant positive effect of the intervention (odds ratio = 0.11; 95%-CI: [0.04; 0.29]).

Anxiety.

Post-treatment. Out of three studies measuring anxiety post-treatment, effect size data were available for two of them. Therefore, these data are synthesized narratively. In the study by Bérubé et al. (2019), 56 subjects being at risk for developing chronic pain after a major lower extremity trauma were randomized either to a self-management intervention or to SC/TAU. The effect of the intervention was not statistically significant (g = -0.19, 95%-CI: [-0.75, 0.37]).

In the study by Sterling, Smeets, Keijzers, Warren, Kenardy (2019), patients suffering from acute whiplash-associated disorder (n = 108) received either stress inoculation training and exercise or exercise alone (SC/TAU). There was no significant effect of the intervention on anxiety post-treatment (g = 0.95%-CI: [-0.39, 0.39]).

Follow-up. Out of six studies measuring anxiety at follow-up, effect size data were available for five studies (n = 481). Follow-up length ranged from 1.5 months to 12 months (*Median* = 10.5). There was a small negative and non-significant effect ($g = -0.018$, 95%-CI: [-0.24, 0.2], see Figure H11). Heterogeneity was not significantly different from zero ($Q(4)$ = 2.92, $p = .57$) and inconsistency was small to considerable ($I^2 = 2.94\%$, 95%-CI: [0%; 78.3%]). The resulting 95%-prediction interval ranged from -0.26 to 0.22.

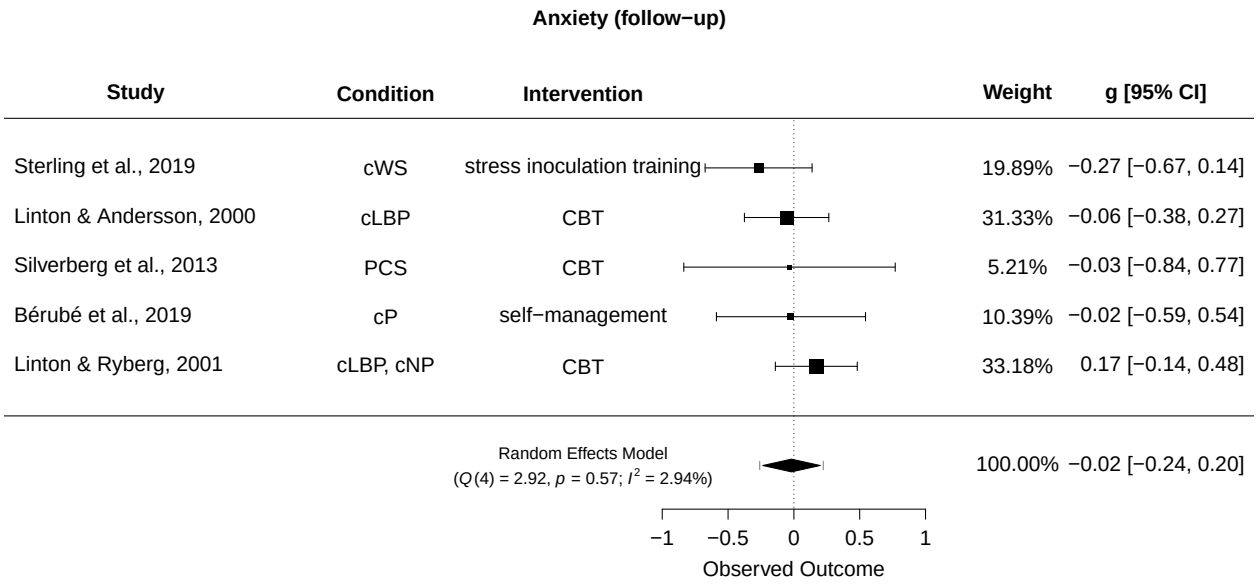

Figure H11. Forest plot of anxiety (follow-up). $g > 0$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. cP: Chronic pain. cWS: Chronic whiplash syndrome. PCS: Post-concussion syndrome.

Risk of bias in individual studies. The risk of bias ratings for each domain are depicted in Figure H12.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.12$ (95%-CI: [-0.79, 1.02]). The 3PSM revealed a corrected effect estimate of g = -0.006 (95%-CI: [-0.2, 0.19]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 0.22$, p = .64).

Additional analyses. No moderator analysis of intervention intensity could be conducted as all studies examined high intensity interventions. No moderator analysis of mean symptom duration could be conducted as there were no observations. Type of population did not significantly moderate the treatment effect ($F(1,3) = 0.0004$, p = .99, R^2 = 0%). No moderator analysis of type of control group could be conducted as all studies examined SC/TAU controls. Length of follow-up did not significantly moderate the treatment effect ($F(1,3) = 0.072$, p = .81, R^2 = 0%).

There was a large positive rank correlation between type of population and length of

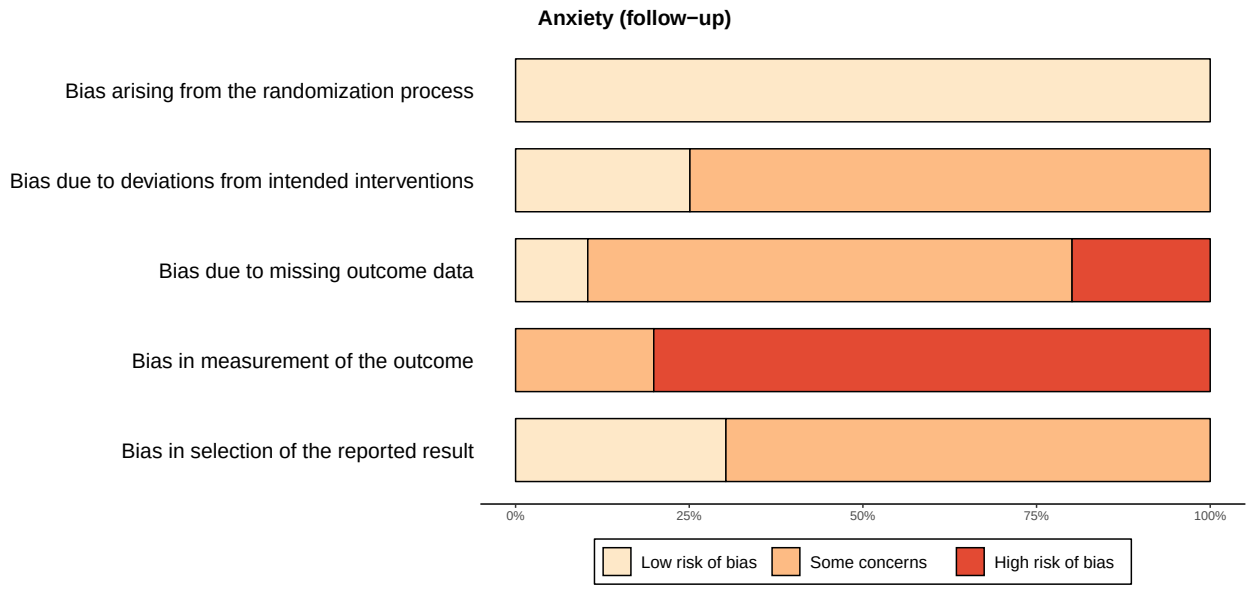

Figure H12. Risk of bias inherent in the summary effect for anxiety (follow-up). Study-level biases are weighted according to the meta-analytic weights.

follow-up ($\rho = .89$).

Depression.

Post-treatment. Out of six studies measuring depression post-treatment, effect size data were available for five studies (n = 720). There was a small significant effect ($g = 0.12$, 95%-CI: [0.03, 0.2], see Figure H13). Heterogeneity was not significantly different from zero ($Q(4) = 0.64$, $p = .96$) and inconsistency was small ($I^2 = 0\%$, 95%-CI: [0%, 24%]). The resulting 95%-prediction interval ranged from 0.03 to 0.2.

Risk of bias in individual studies. For a summary of risk of bias ratings, see Figure H14.

Meta-bias. The PET revealed a corrected effect estimate of $g = 0.12$ (95%-CI: [-0.4, 0.64]). The 3PSM revealed a corrected effect estimate of $g = 0.17$ (95%-CI: [0.046, 0.29]. A likelihood-ratio test did not reveal a significantly better fit of the bias-adjusted model to the data ($\chi^2(1) = 1.32$, $p = .25$).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,3) = 1.34$, $p = .33$, $R^2 = 0\%$). No moderator analysis of mean symptom duration could be computed as there were too few observations ($k = 2$). Type of population did not significantly moderate the treatment effect ($F(1,3) = 0.22$, $p = .67$, $R^2 = 0\%$). Type of control

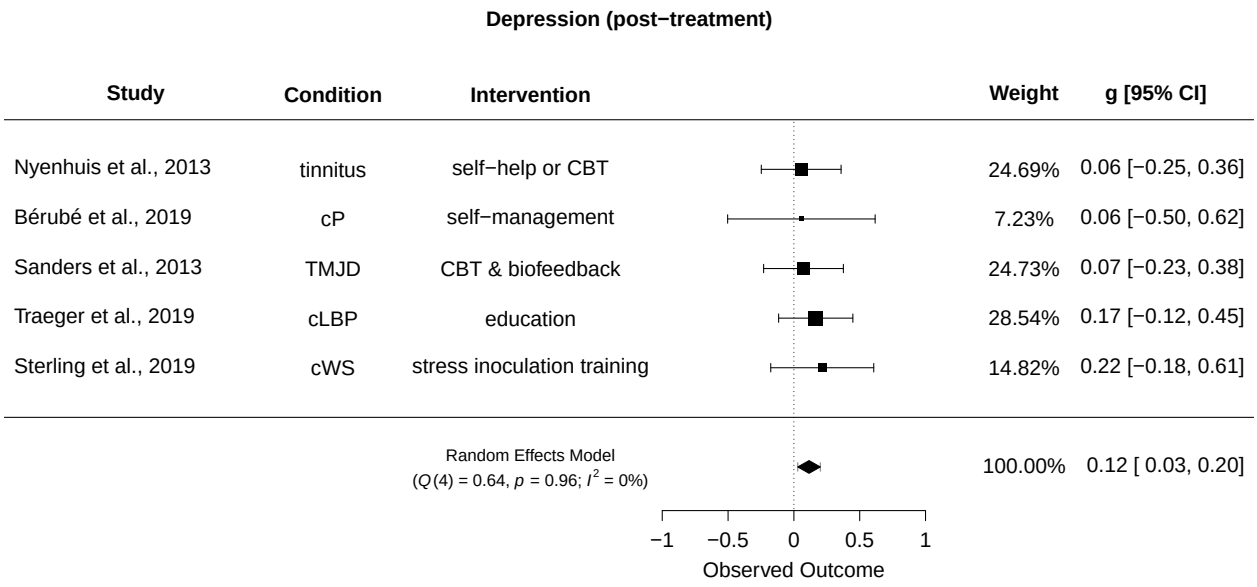

Figure H13. Forest plot of depression (post-treatment). $g > 1$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cP: Chronic pain. cWS: Chronic whiplash syndrome. TMJD: Temporomandibular joint disorder.

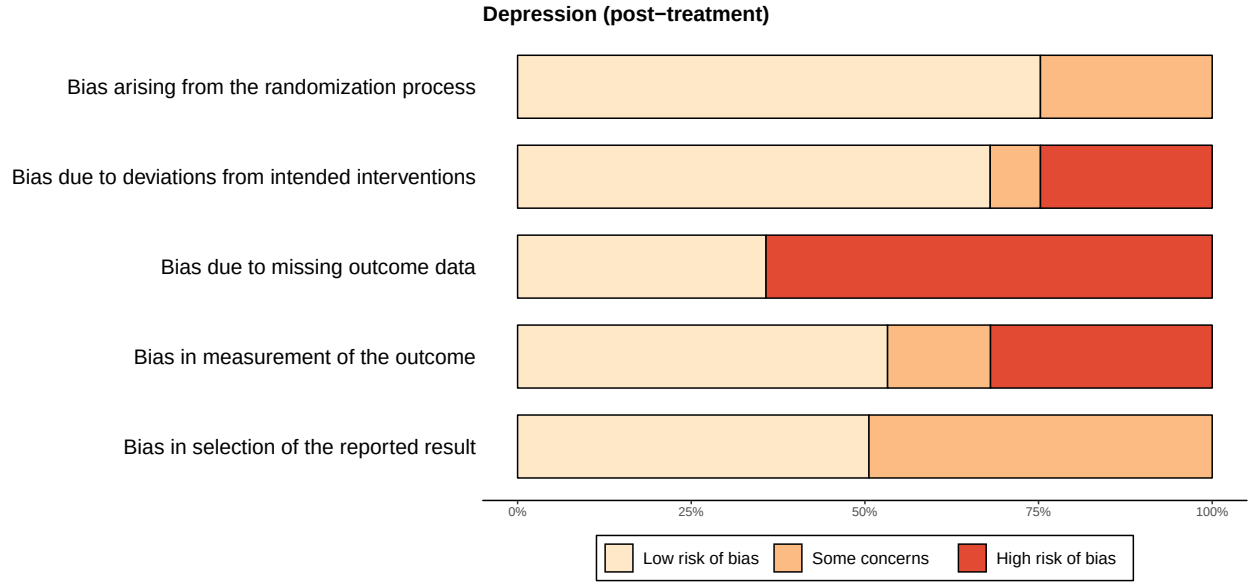

Figure H14. Risk of bias inherent in the summary effect for depression (post-treatment). Study-level biases are weighted according to the meta-analytic weights.

group did not significantly moderate the treatment effect ($F(2,2) = 0.6$, $p = .62$, $R^2 = 0\%$).

There was a small interdependence between intervention intensity and type of population ($V = .25$) with all studies with prevention populations investigating high intensity interventions. There was a perfect interdependence between intervention intensity and type of control group ($V = 1$) with high intensity interventions being only investigated in studies with SC/TAU or placebo controls and low intensity interventions being investigated in studies with no treatment controls, only. There was a large interdependence between type of population and type of comparison ($V = .61$) with no treatment and placebo controls being only employed in studies with early intervention populations.

Follow-up. Out of 10 studies measuring depression at follow-up, effect size data were available for nine studies ($n = 1063$). Follow-up length ranged from 1.5 months to 12 months (*Median* = 9.5). There was a small and non-significant effect ($g = 0.1$, 95%-CI: [-0.016, 0.21], see Figure H15). Heterogeneity was not significantly different from zero ($Q(8) = 4.83$, $p = .78$) and inconsistency was small to substantial ($I^2 = 0.015\%$, 95%-CI: [0%, 70.5%]). The resulting 95%-prediction interval ranged from -0.017 to 0.21.

Risk of bias in individual studies. Figure H16 depicts the risk of bias ratings.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = -0.046$ (95%-CI: [-0.19, 0.097]). The 3PSM revealed a corrected effect estimate of $g = 0.14$. A confidence interval could not be computed due to problems with model convergence. A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 1.65$, $p = .2$).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,7) = 0.67$, $p = .44$, $R^2 = 99.4\%$). Mean symptom duration did not significantly moderate the treatment effect ($F(1,1) = 0.037$, $p = .88$, $R^2 = 0\%$). Type of population did not significantly moderate the treatment effect ($F(1,7) = 2.83$, $p = .14$, $R^2 = 100\%$). Type of control group did not significantly moderate the treatment effect ($F(2,6) = 0.086$, $p = .92$, $R^2 = 52.4\%$). Length of follow-up did not significantly moderate the treatment effect ($F(1,7) = 0.84$, $p = .39$, $R^2 = 100\%$).

There was a medium-sized negative correlation between intervention intensity and mean

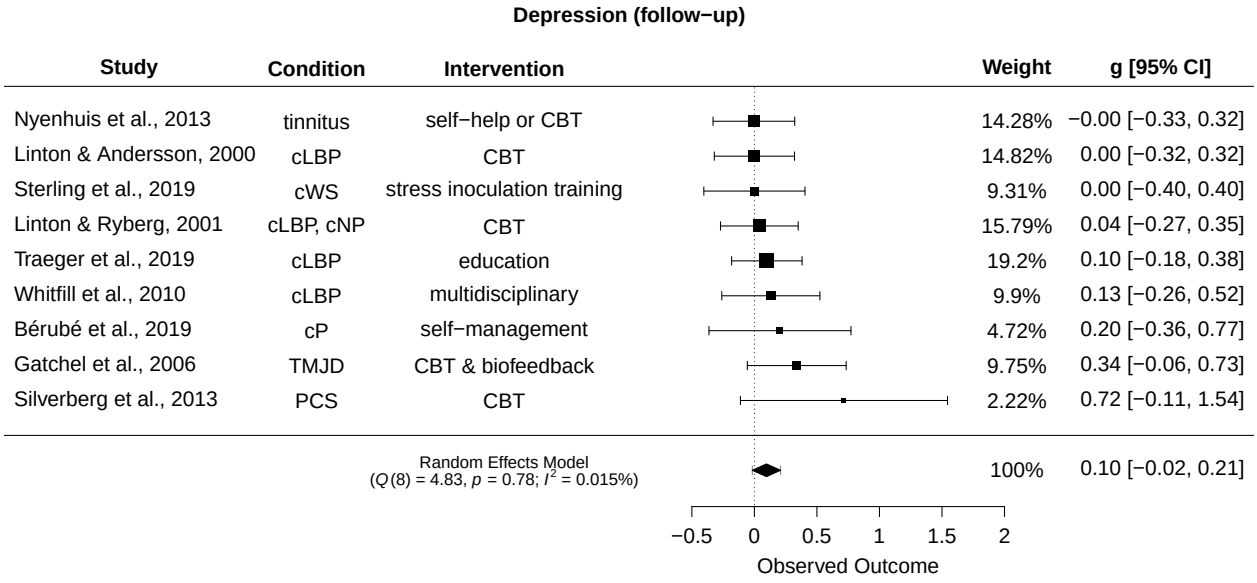

Figure H15. Forest plot of depression (follow-up). $g > 0$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. cP: Chronic pain. cWS: Chronic whiplash syndrome. PCS: Post-concussion syndrome. TMJD: Temporomandibular joint disorder

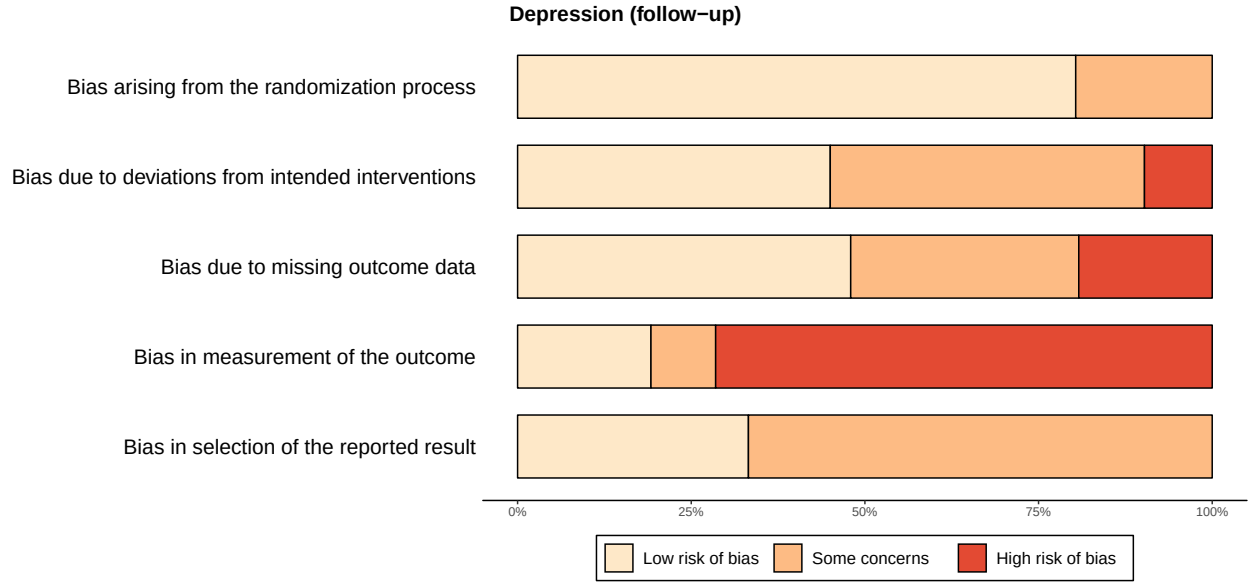

Figure H16. Risk of bias inherent in the summary effect for depression (follow-up). Study-level biases are weighted according to the meta-analytic weights.

symptom duration ($r_b = -.49$). There was a small interdependence between intervention intensity and type of population ($V = .19$) with all studies with prevention populations investigating high intensity interventions. There was a large interdependence between intervention intensity and type of control group ($V = .66$) with all SC/TAU and placebo comparisons being conducted in studies evaluating high intensity interventions. There was a small negative correlation between intervention intensity and length of follow-up ($r_b = -.11$). No rank correlation between mean symptom duration and type of population could be computed since all studies providing mean symptom duration data were conducted in early intervention populations. There was a large negative rank correlation between symptom duration and type of comparison ($\rho = -.87$). There was a nearly perfect positive correlation between length of follow-up and symptom duration ($r = .99$). There was a medium-sized interdependence between type of population and type of control group ($V = .38$) with all no treatment and placebo comparisons being conducted in studies with early intervention populations. There was a large positive rank correlation between type of population and length of follow-up ($\rho = .58$). There was a small negative rank correlation between type of comparison and length of follow-up ($\rho = -.29$).

Health care utilization.

Post-treatment. Out of four studies measuring health care utilization post-treatment, effect size data were available for none of them.

Follow-up. Out of eight studies measuring health care utilization at follow-up, effect size data were available for three studies (n = 283). Follow-up length ranged from 1.5 months to 12 months (*Median* = 10.5). There was a positive small and significant effect ($g = 0.31$, 95%-CI: [0.18, 0.44], see Figure H17). Heterogeneity was not significantly different from zero ($Q(2) = 0.13$, $p = .94$) and inconsistency was small to substantial ($I^2 = 0\%$, 95%-CI: [0%, 76.4%]). The resulting 95%-prediction interval ranged from 0.18 to 0.44.

Risk of bias in individual studies. Figure H18 summarizes the risk of bias ratings.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.26$ (95%-CI: [0.15, 0.38]). The 3PSM revealed a corrected effect estimate of $g = 0.82$. A confidence interval could not be computed due to problems with model convergence. A likelihood-ratio

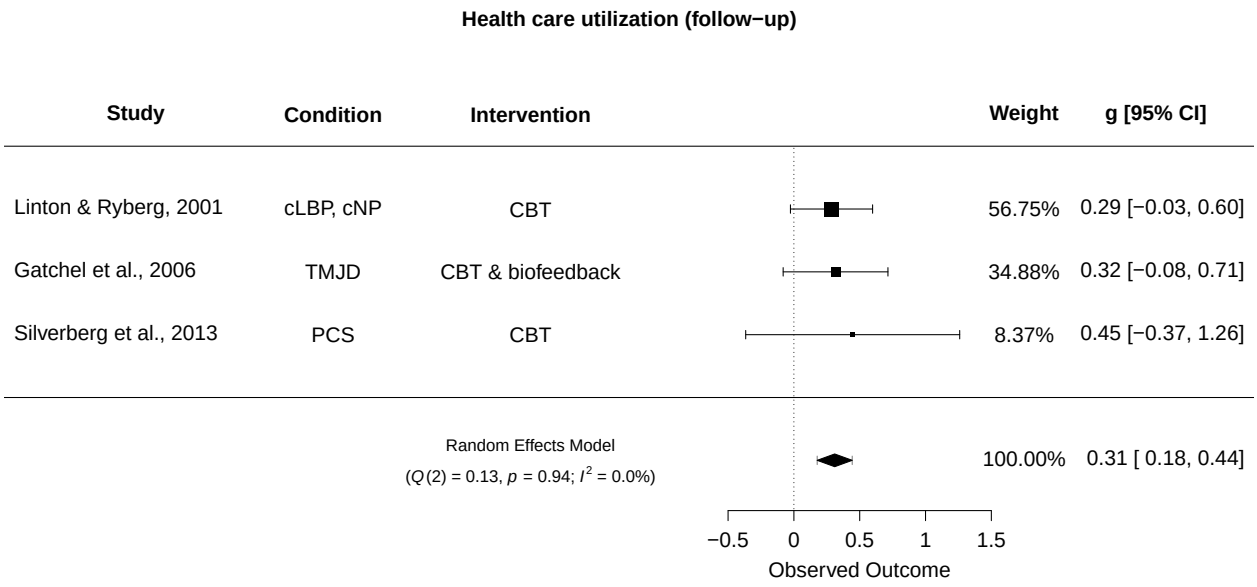

Figure H17. Forest plot of health care utilization (follow-up). $g > 1$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. PCS: Post-concussion syndrome. TMJD: Temporomandibular joint disorder.

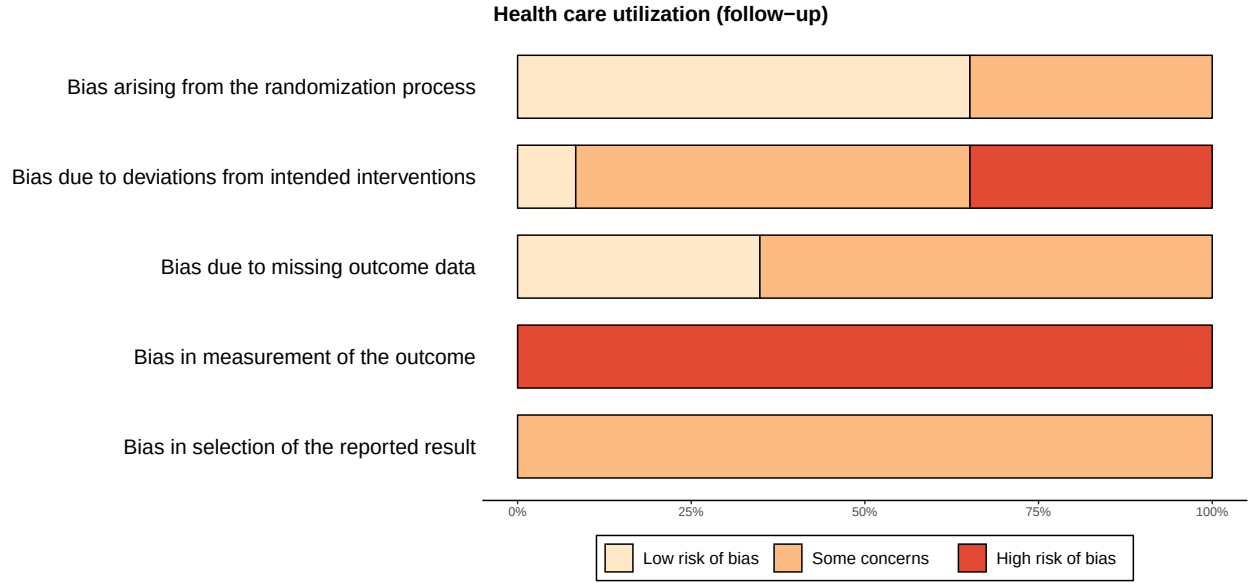

Figure H18. Risk of bias inherent in the summary effect for health care utilization (follow-up). Study-level biases are weighted according to the meta-analytic weights.

test revealed a significantly better fit of the 3PSM to the data ($\chi^2(1) = 5.75$, $p = .016$).

Additional analyses. No moderator analysis of intervention intensity could be computed as all studies evaluated high intensity interventions. No moderator analysis of mean symptom duration could be computed as there were too few available studies ($k = 1$). Type of population did not significantly moderate the treatment effect ($F(1,1) = 8.46$, $p = .21$, $R^2 = 0\%$). Type of control group did not significantly moderate the treatment effect ($F(1,1) = 0.012$, $p = .93$, $R^2 = 0\%$). Length of follow-up did not significantly moderate the treatment effect ($F(1,1) = 145.3$, $p = .053$, $R^2 = 0\%$).

There was a large interdependence between type of population and type of control group ($V = .5$) with all no treatment comparisons being conducted in studies with early intervention populations. There was a large positive rank correlation between type of population and length of follow-up ($\rho = .87$). There was a no rank correlation between type of comparison and length of follow-up ($\rho = 0$).

Consumer satisfaction.

Post-treatment. Out of six studies measuring consumer satisfaction post-treatment, appropriate effect size data were available for two studies ($n = 371$). Therefore, the data were synthesized narratively. In the study by Damush et al. (2003b), subjects with acute low back pain participated in a self-management program while control subjects received SC/TAU. Based on a sample of $n = 163$, there was no significant effect of the intervention ($g = -0.02$, 95%-CI: [-0.33; 0.29]).

In the study by Nyenhuis, Zastrutzki, Weise, et al. (2013), subjects suffering from acute tinnitus were treated either with group CBT, bibliotherapy or an online self-help program in the intervention groups. Except for an information sheet concerning the auditory system, tinnitus and treatment options, the control subjects received no treatment. There was a large significant combined effect of the interventions ($g = 1.21$, 95%-CI: [0.89; 1.54], $n = 208$).

Although the other studies did not provide appropriate data for meta-analytic integration, there was other information concerning the consumer satisfaction available. In the study by Gil-Jardiné et al. (2018) evaluating EMDR or reassurance compared to SC/TAU in patients at high risk for post-concussion syndrome, consumer satisfaction was rated on an

11-point numeric rating scale ranging from 0 to 10 with higher values indicating higher satisfaction. There was a median satisfaction of 9.5 (interquartile range (*IQR*): 8 - 10, *n* = 34) in the EMDR group, a median satisfaction of 8.5 (*IQR*: 7.25 - 10, *n* = 38) in the reassurance group and a median satisfaction of 8 (*IQR*: 6 - 10, *n* = 37) in the control group.

In the study by Karjalainen et al. (2004), consumer satisfaction was rated on the same scale. Subjects were patients suffering from subacute low back pain. The intervention consisted of advice, physiotherapeutic exercises and for a subset of subjects also of a worksite visit by a physiotherapist and a physician. The control group received SC/TAU. Intervention groups resulted in a combined mean satisfaction of 6.15 (range: 0 - 10, *n* = 104), while the SC/TAU group resulted in a mean satisfaction of 4.1 (range: 0 - 10, *n* = 56).

Follow-up. Out of five studies measuring consumer satisfaction at follow-up, effect size data were available for one study. In the study by Damush et al. (2003b, described above), there was no significant difference between the intervention and the control group ($g = 0.098$, 95%-CI: [-0.24; 0.43], *n* = 139) after a follow-up length of 11.25 months.

There were two further studies with relevant data, although they did not report enough data for calculating an effect size. In the study by Karjalainen et al. (2004, described above) there was a combined mean satisfaction of 5.99 (range: 0 - 10, *n* = 103) in the intervention groups and a mean satisfaction of 4.3 (range: 0 - 10, *n* = 53) in the SC/TAU group after at the 24-months follow-up.

In the study by Silverberg et al. (2013), subjects at risk for post-concussion syndrome received six sessions of CBT. Consumer satisfaction was assessed using a 5-point Likert scale. At 1.5 months follow-up, the mean satisfaction in the intervention group was 4.69 (*SD* = 0.48, *n* = 13) indicating high satisfaction. There were no data available for the SC/TAU control group.

Online Supplement I

Sensitivity analyses: two-step DerSimonian-Laird estimator

Primary outcomes

Somatic symptom severity.

Post-treatment. Out of 19 studies measuring somatic symptom severity post-treatment, effect size data were available for 13 studies (n = 2,031). There was a small and non-significant effect ($g = 0.1$, 95%-CI: [-0.079, 0.3], see Figure I1). Heterogeneity was significantly different from zero ($Q(12) = 38.3$, $p = .0001$) and inconsistency was moderate to considerable ($I^2 = 66.5\%$, 95%-CI: [33.3%, 88.8%]). The resulting 95%-prediction interval ranged from -0.45 to 0.67.

Risk of bias in individual studies. Risk of bias inherent in the summary effect is depicted in Figure I2.

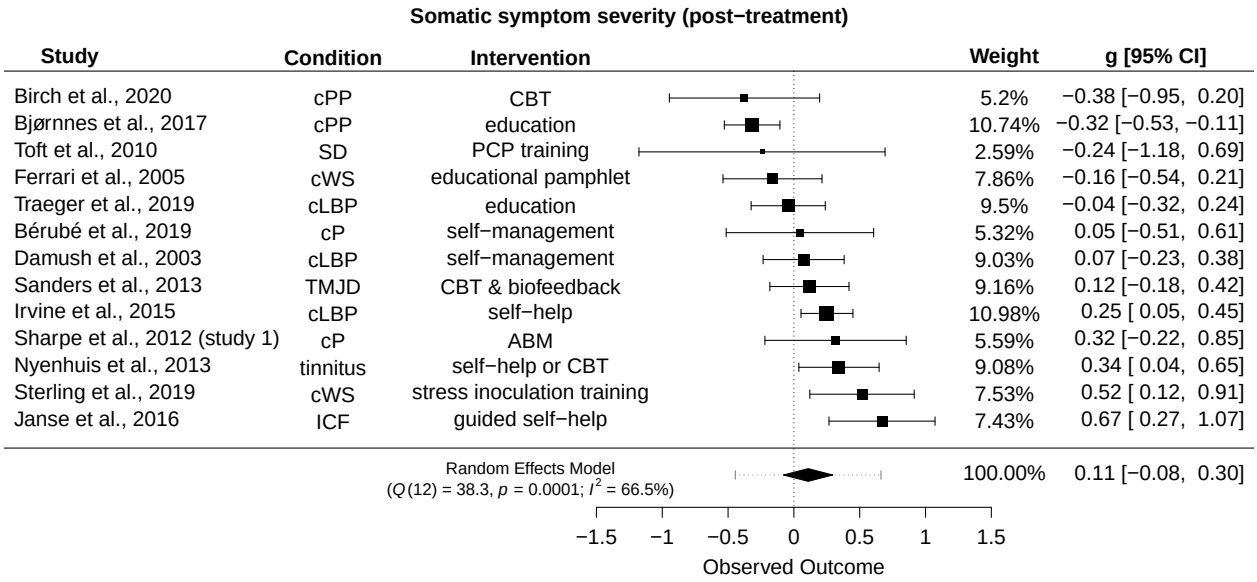

Figure I1. Forest plot of somatic symptom severity (post-treatment). $g > 0$ indicates more favorable outcomes in the intervention group. ABM: Attention bias modification. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cP: Chronic pain. cPP: Chronic postoperative pain. cWS: Chronic whiplash syndrome. ICF: Idiopathic chronic fatigue. PCP: Primary care physician. SD: Somatoform disorder. TMJD: Temporomandibular joint disorder.

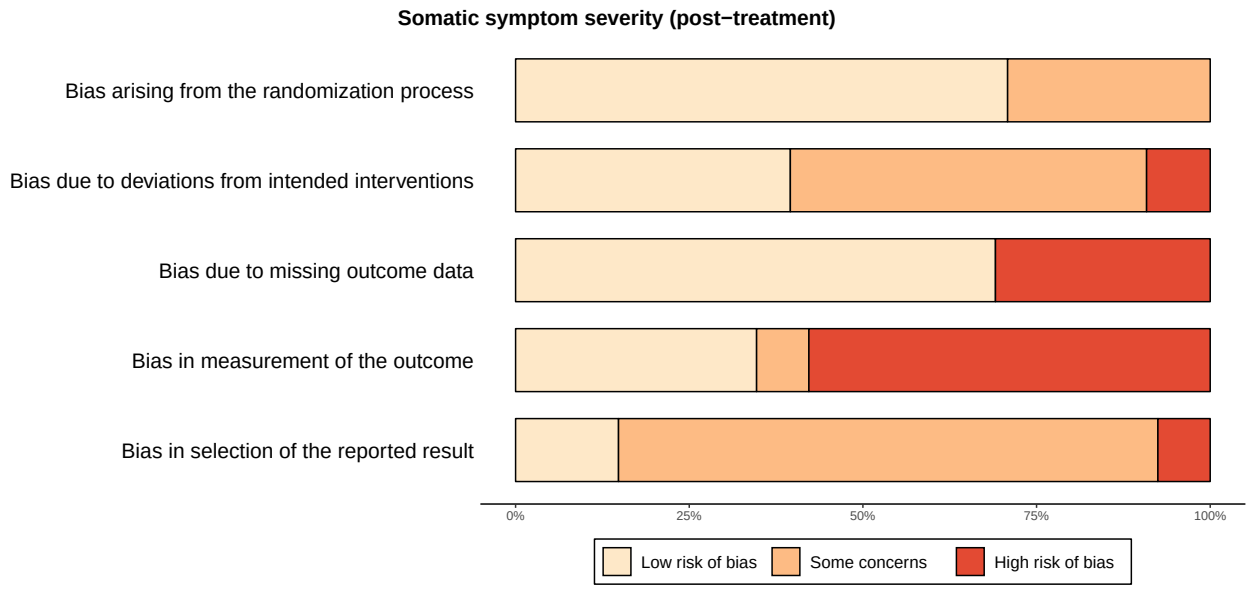

Figure I2. Risk of bias inherent in the summary effect for somatic symptom severity (post-treatment). Study-level biases are weighted according to the meta-analytic weights. One cluster-randomized study was included in this meta-analysis (Toft et al., 2010). There was a high risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization in this study (not depicted).

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.008$ (95%-CI: [-0.51, 0.52]). The 3PSM revealed a corrected effect estimate of $g = -0.023$ (95%-CI: [-0.2, 0.15]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 2.68$, $p = .1$).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,11) = 0.16$, $p = .7$, $R^2 = 0\%$). No moderator analysis of mean symptom duration could be computed as there were too few observations ($k = 2$). Type of population significantly moderated the treatment effect ($F(1,11) = 6.85$, $p = .024$, $R^2 = 54.9\%$). Specifically, there was no significant effect in studies with prevention populations ($g = -0.15$, 95%-CI: [-0.42, 0.11]), while there was a significant effect for studies with early intervention populations ($g = 0.23$, 95%-CI: [0.047, 0.42]). Type of control group did not significantly moderate the treatment effect ($F(3,9) = 2.84$, $p = .098$, $R^2 = 52.8\%$).

Descriptive analyses revealed a medium-sized interdependence between intervention

intensity and type of population ($V = .35$) resulting from high intensity interventions being over-represented in studies with early intervention populations. There was a large interdependence between intervention intensity and type of control group ($V = .56$) with all no treatment comparisons being conducted in studies evaluating low intensity interventions and all wait-list comparisons being conducted in studies evaluating high intensity interventions. There was a medium-sized interdependence between type of population and type of control group ($V = .48$) with all no treatment and wait-list comparisons being conducted in studies with early intervention populations.

Follow-up. Out of 24 studies measuring somatic symptom severity at follow-up, effect size data were available for 17 studies (n = 2,438). Follow-up length ranged from 1.5 months to 24 months (*Median* = 9.5). There was a small and significant positive effect ($g = 0.25$, 95%-CI: [0.097, 0.41], see Figure I3). Heterogeneity was significantly different from zero ($Q(16) = 37.4$, $p = .002$) and inconsistency was small to considerable ($I^2 = 61.5\%$, 95%-CI: [22.9%, 85.9%]). The resulting 95%-prediction interval ranged from -0.24 to 0.75.

Risk of bias in individual studies. Risk of bias inherent in the summary effect is depicted in Figure I4.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.04$ (95%-CI: [-0.32, 0.4]). The 3PSM revealed a corrected effect estimate of $g = 0.14$ (95%-CI: [-0.041, 0.31]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 2.01$, p = .16).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,15) = 0.031$, p = .86, R^2 = 0%). Mean symptom duration did not significantly moderate the treatment effect ($F(1,1) = 1.04$, p = .49, R^2 = 0%). Type of population did not significantly moderate the treatment effect ($F(1,15) = 1.37$, p = .26, R^2 = 0%). Type of control group did not significantly moderate the treatment effect ($F(2,14) = 0.54$, p = .6, R^2 = 0%). Length of follow-up did not significantly moderate the treatment effect ($F(1,15) = 0.43$, p = .52, R^2 = 0%).

There was a medium-sized negative correlation between intervention intensity and mean symptom duration ($r_b = -.49$) with high intensity interventions displaying lower mean

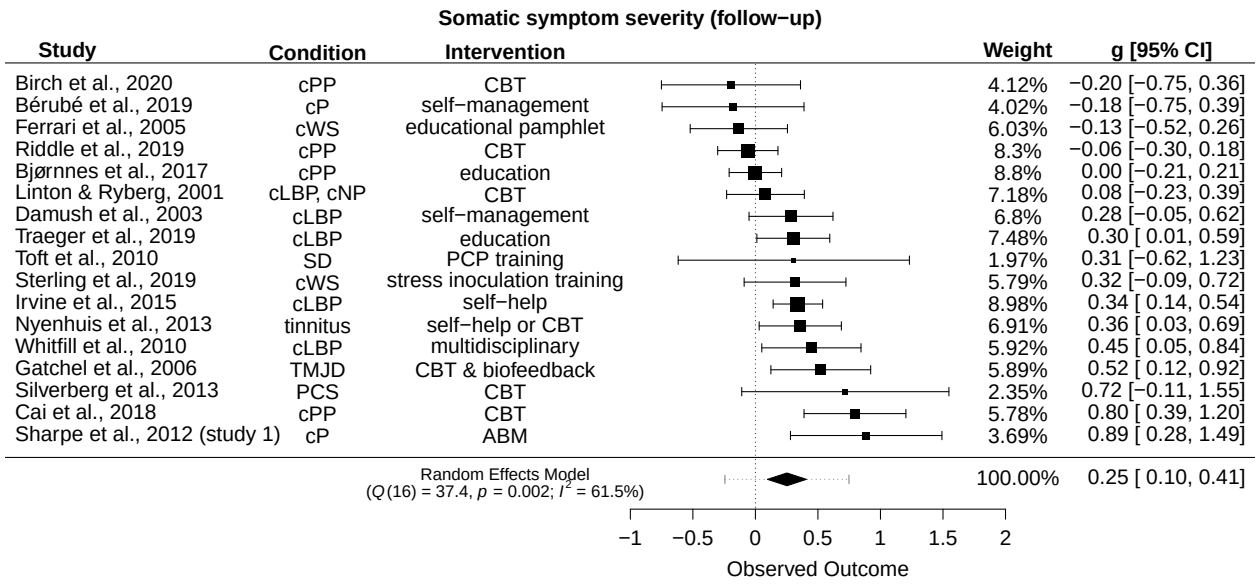

Figure I3. Forest plot of somatic symptom severity (follow-up). $g > 0$ indicates more favorable outcomes in the intervention group. ABM: Attention bias modification. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. cP: Chronic pain. cPP: Chronic postoperative pain. cWS: Chronic whiplash syndrome. PCP: Primary care physician. PCS: Post-concussion syndrome. SD: Somatoform disorder. TMJD: Temporomandibular joint disorder.

symptom durations. There was a small interdependence between intervention intensity and type of population ($V = .17$) resulting from high intensity interventions being over-represented in studies with early intervention populations. There was a medium-sized interdependence between intervention intensity and type of control group ($V = .4$) with high intensity interventions being over-represented in studies with SC/TAU controls. There was a medium-sized correlation between intervention intensity and length of follow-up ($r_b = .35$) with high intensity interventions displaying bigger lengths of follow-up. No rank correlation between mean symptom duration and type of population could be computed as all studies providing mean symptom duration data were conducted in early intervention populations. There was a large negative rank correlation between mean symptom duration and type of control group ($\rho = -.87$). There was a medium-sized negative correlation between mean symptom duration and length of follow-up ($r = -.49$). There was a medium-sized

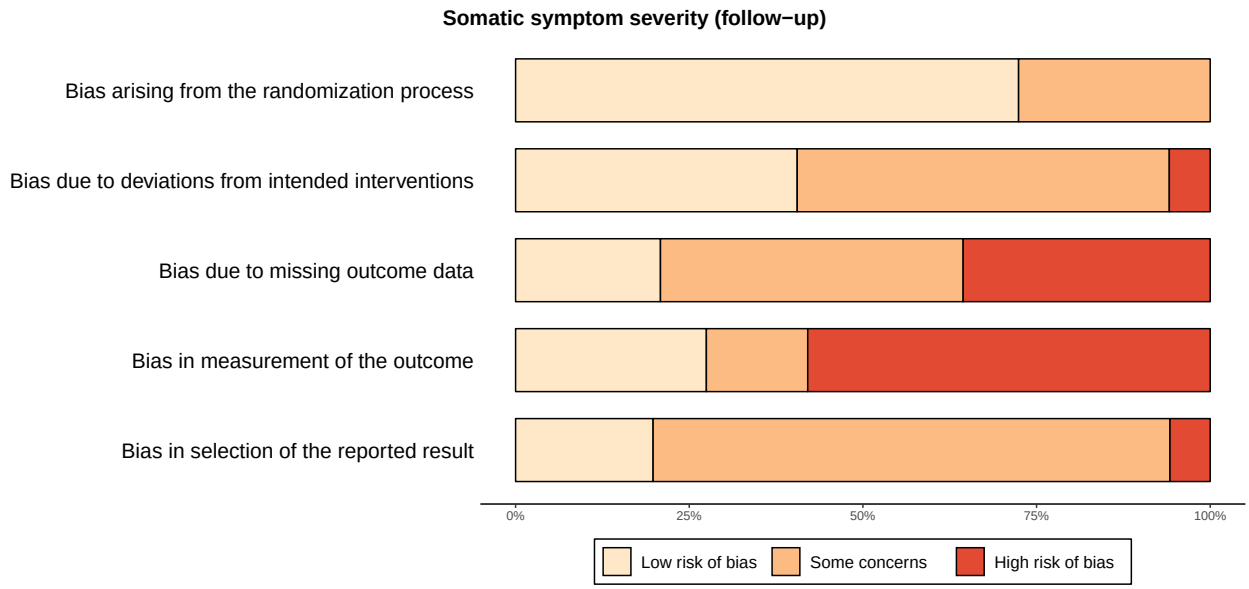

Figure I4. Risk of bias inherent in the summary effect for somatic symptom severity (follow-up). Study-level biases are weighted according to the meta-analytic weights. One cluster-randomized study was included in this meta-analysis (Toft et al., 2010). There was a high risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization in this study (not depicted).

interdependence between type of population and type of control group ($V = .45$) with all no treatment comparisons being conducted in studies with early intervention populations. There was a medium-sized positive rank correlation between type of population and length of follow-up ($\rho = .45$). There was a small positive rank correlation between type of comparison and length of follow-up ($\rho = .12$).

Health-related quality of life.

Post-treatment. Out of 17 studies measuring health-related quality of life post-treatment, effect size data were available for 11 studies (n = 4,498). There was a small and non-significant effect ($g = 0.13$, 95%-CI: [-0.077, 0.33], see Figure I5). Heterogeneity was significantly different from zero ($Q(10) = 19.9$, $p = .03$) and inconsistency was small to considerable ($I^2 = 56.9\%$, 95%-CI: [0%, 88.8%]). The resulting 95%-prediction interval ranged from -0.39 to 0.64.

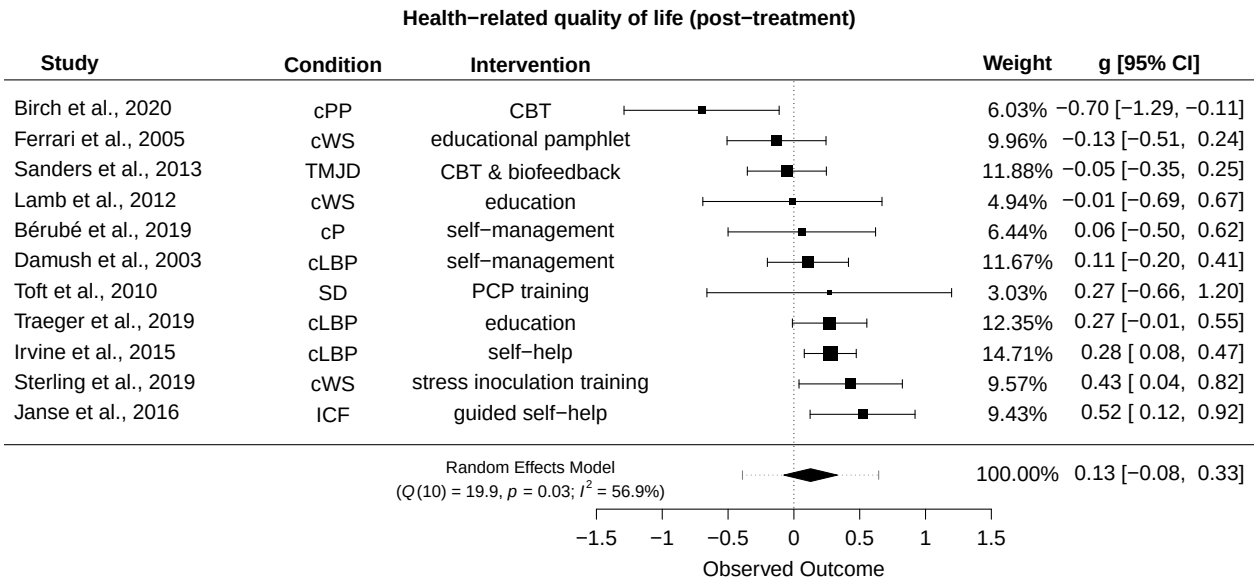

Figure I5. Forest plot of health-related quality of life (post-treatment). $g > 0$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cP: Chronic pain. cPP: Chronic postoperative pain. cWS: Chronic whiplash syndrome. ICF: Idiopathic chronic fatigue. PCP: Primary care physician. SD: Somatoform disorder. TMJD: Temporomandibular joint disorder.

Risk of bias in individual studies. Risk of bias inherent in the summary effect is depicted in Figure I6.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.23$ (95%-CI: [-0.013, 0.48]). The 3PSM revealed a corrected effect estimate of $g = 0.055$ (95%-CI: [-0.13, 0.24]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 1.79$, $p = .18$).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,9) = 0.065$, $p = .81$, $R^2 = 0\%$). No moderator analysis of mean symptom duration could be conducted as there were too few observations ($k = 1$). Type of population significantly moderated the treatment effect ($F(1,9) = 6.99$, $p = .027$, $R^2 = 71.8\%$). Specifically, there was no significant effect in studies with prevention populations ($g = -0.19$, 95%-CI: [-0.51, 0.14]), while there was a significant effect for studies with early intervention populations ($g = 0.24$, 95%-CI: [0.071, 0.41]). Type of control group did not significantly

Figure I6. Risk of bias inherent in the summary effect for health-related quality of life (post-treatment). Study-level biases are weighted according to the meta-analytic weights. Two cluster-randomized studies were included in this meta-analysis (Lamb et al., 2012; Toft et al., 2010). While the study by Lamb et al. (2012) was at low risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization, the study by Toft et al. (2010) was at high risk (not depicted).

moderate the treatment effect ($F(3,7) = 0.95$, $p = .47$, $R^2 = 0\%$).

There was a medium-sized interdependence between intervention intensity and type of population ($V = .39$) resulting from high intensity interventions being over-represented in studies with early intervention populations. There was a large interdependence between intervention intensity and type of control group ($V = .59$) with all no treatment comparisons being conducted in studies evaluating low intensity interventions and all wait-list and placebo comparisons being conducted in studies evaluating high intensity interventions. There was a large interdependence between type of population and type of control group ($V = .57$) with prevention populations being used in studies having SC/TAU controls, only.

Follow-up. Out of 18 studies measuring health-related quality of life at follow-up, effect size data were available for 12 studies (n = 1,681). Follow-up length ranged from 2 months to 24 months (*Median* = 10). There was a positive small and significant effect

($g = 0.13$, 95%-CI: [0.007, 0.25], see Figure I7). Heterogeneity was not significantly different from zero ($Q(11) = 13.2$, $p = .28$) and inconsistency was small to substantial ($I^2 = 11.4\%$, 95%-CI: [0%, 72.7%]). The resulting 95%-prediction interval ranged from -0.058 to 0.31.

Risk of bias in individual studies. Risk of bias is depicted in Figure I8.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.18$ (95%-CI: [0.002, 0.36]). The 3PSM revealed a corrected effect estimate of $g = 0.16$ (95%-CI: [-0.004, 0.32]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 0.45$, $p = .5$).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,10) = 3.68$, $p = .084$, $R^2 = 100\%$). No moderator analysis of mean symptom duration could be conducted as there were too few observations ($k = 1$). Type of population significantly moderated the treatment effect ($F(1,10) = 6.14$, $p = .033$, $R^2 = 100\%$). Specifically, there was no significant effect in studies with prevention populations

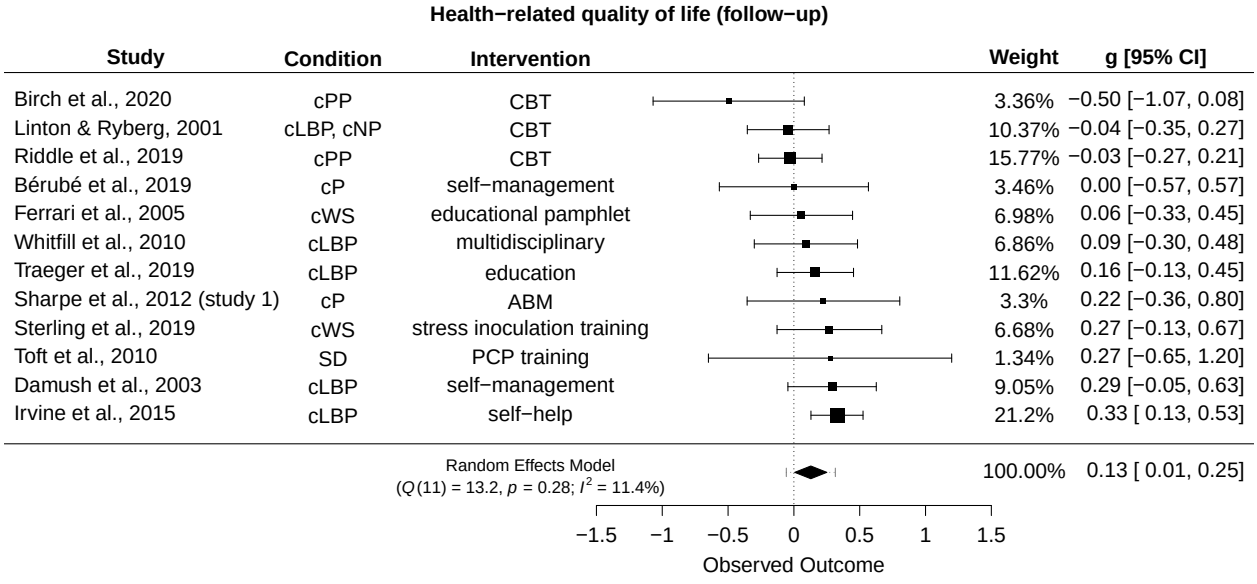

Figure I7. Forest plot of health-related quality of life (follow-up). $g > 0$ indicates more favorable outcomes in the intervention group. ABM: Attention bias modification. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cP: Chronic pain. cPP: Chronic postoperative pain. cWS: Chronic whiplash syndrome. PCP: Primary care physician. SD: Somatoform disorder.

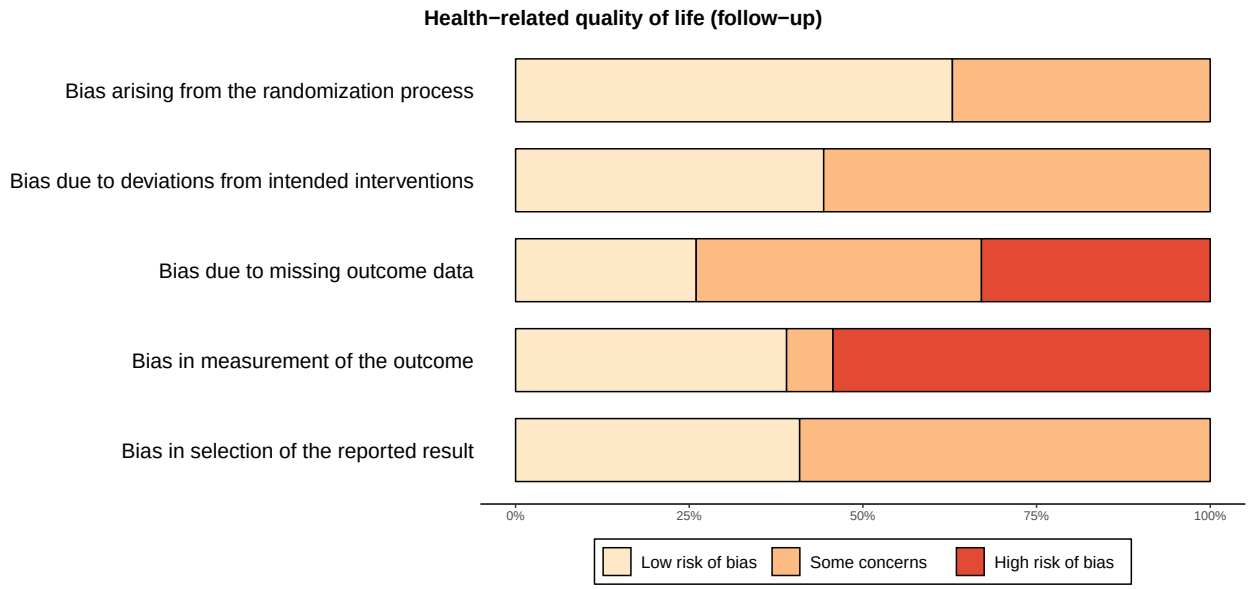

Figure I8. Risk of bias inherent in the summary effect for health-related quality of life (follow-up). Study-level biases are weighted according to the meta-analytic weights. One cluster-randomized study was included in this meta-analysis (Toft et al., 2010). There was a high risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization in this study (not depicted).

($g = -0.023$, 95%-CI: [-0.21, 0.15]), while there was a significant effect for studies with early intervention populations ($g = 0.21$, 95%-CI: [0.091, 0.34]). Type of control group did not significantly moderate the treatment effect ($F(2,9) = 2.57$, $p = .13$, $R^2 = 100\%$). Length of follow-up did not significantly moderate the treatment effect ($F(1,10) = 1.24$, $p = .29$, $R^2 = 59\%$).

There was a small interdependence between intervention intensity and type of population ($V = .29$) with high intensity interventions being over-represented in studies with early intervention populations. There was a large interdependence between intervention intensity and type of control group ($V = .56$) with all no treatment comparisons being conducted in studies evaluating low intensity interventions. There was a large correlation between intervention intensity and length of follow-up ($r_b = .63$). There was a medium-sized interdependence between type of population and type of control group ($V = .36$) with all no treatment comparisons being conducted in studies with early intervention populations. There

was a large positive rank correlation between type of population and length of follow-up ($\rho = .52$). There was a small positive rank correlation between type of comparison and length of follow-up ($\rho = .19$).

Secondary outcomes

Unwanted negative treatment effects.

Post-treatment. Since only one study assessed unwanted negative treatment effects post-treatment, we describe these data narratively. Sterling, Smeets, Keijzers, Warren, Kenardy (2019) evaluated the effect of stress inoculation training in combination with guideline-based exercise compared to guideline-based exercise alone (SC/TAU) for patients suffering from whiplash-associated disorder (n = 108). The researchers assessed adverse effects (i.e., exacerbation of a pre-existing condition) and adverse events (i.e., events that are life-threatening, require inpatient hospitalization, or will result in persistent or significant disability or incapacity) via open ended questions. In each trial arm, one subject reported neck pain exacerbation, while no subject reported adverse events.

Follow-up. Only two studies assessed unwanted negative treatment effects at follow-up. Therefore, we describe these data narratively. In the study by Traeger et al. (2019), patients with acute low back pain (n = 202) were randomized to an intensive patient education condition or to a placebo education condition. Both treatments were delivered face-to-face. The researchers recorded adverse events during the trial. Over a follow-up time of 10.5 months, there were no reported adverse events in any of the treatment groups.

In the study by Riddle et al. (2019), patients scheduled for a knee arthroplasty at risk for chronic pain (n = 402) received either CBT-based pain coping skills training or arthritis education serving as placebo condition. Beyond that, there was third trial arm providing SC/TAU, only. Unwanted negative treatment effects were assessed during data collection and by medical record review after a follow-up time of 10.5 months. There were no significant differences neither in adverse events (e.g., emergency room visits due to knee pain, psychological distress, elevated depressive symptoms) nor in serious adverse events (e.g., hospitalization, surgery, infection, death) between groups.

Diagnostic status concerning SSD/FSS.

Post-treatment. Four studies measured diagnostic status concerning SSD/FSS post-treatment. Effect size data were available for all of them (n = 427). A random-effects meta-analysis revealed a risk ratio of 0.92 (95%-CI: [0.62, 1.37], see Figure I9). Heterogeneity was not significantly different from zero ($Q(3) = 7.53$, $p = .057$) and inconsistency was small to considerable ($I^2 = 61.4\%$, 95%-CI: [0%, 97.5%]). The resulting 95%-prediction interval ranged from 0.44 to 1.92.

Risk of bias in individual studies. Figure I10 depicts the risk of bias inherent in the diagnostic status summary effect.

Meta-bias. The PET-PEESE revealed a corrected risk ratio of 0.68 (95%-CI: [0.12, 3.83]). The 3PSM revealed a corrected risk ratio of 0.96 (95%-CI: [0.66, 1.42]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 0.4$, p = .53).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,2) = 1.9$, p = .3, R^2 = 36.2%). No moderator analysis of mean symptom

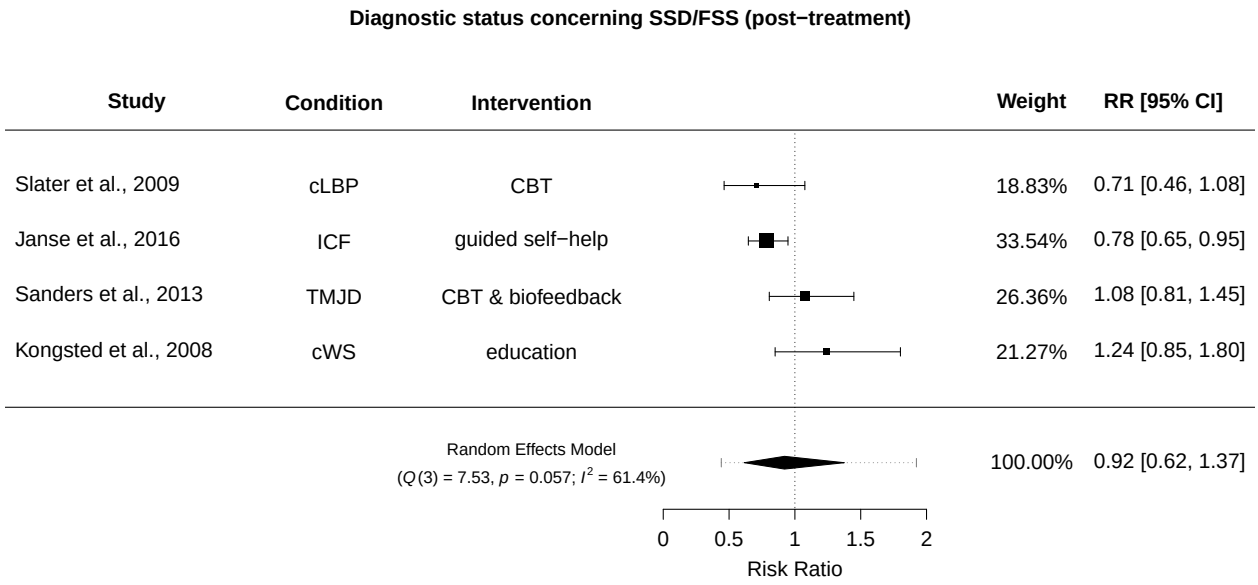

Figure I9. Forest plot of diagnostic status concerning SSD/FSS (post-treatment). *RR* < 1 indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cWS: Chronic whiplash syndrome. ICF: Idiopathic chronic fatigue. TMJD: Temporomandibular joint disorder.

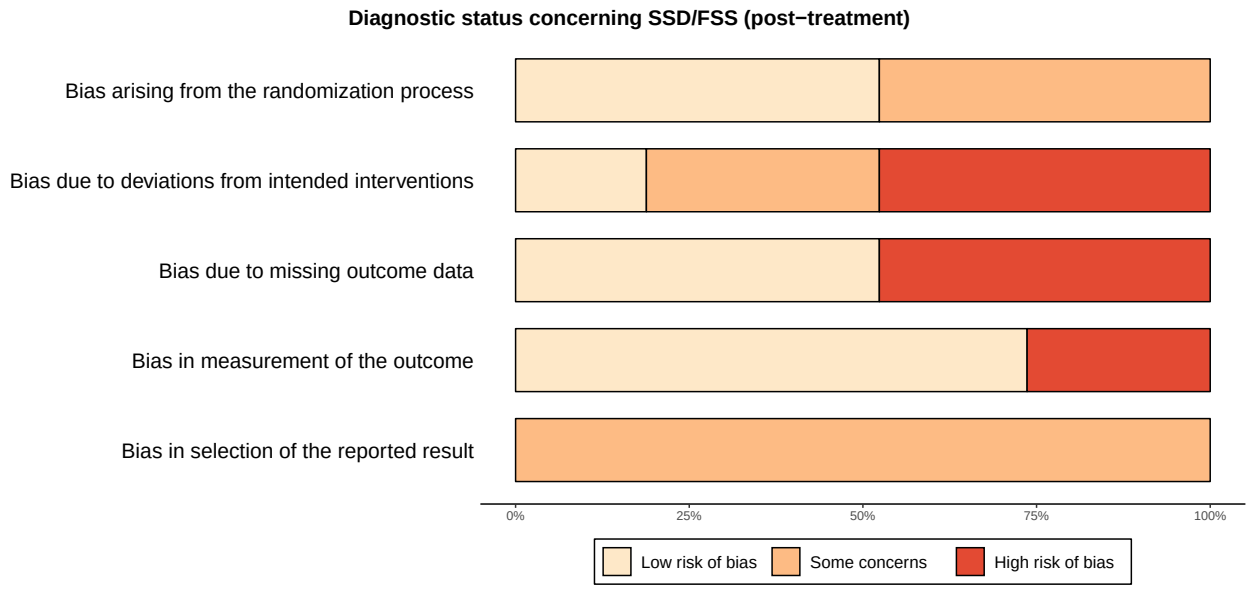

Figure 110. Risk of bias inherent in the summary effect for diagnostic status concerning SSD/FSS (post-treatment). Study-level biases are weighted according to the meta-analytic weights.

duration could be computed as there were no observations. Type of population did not significantly moderate the treatment effect ($F(1,2) = 1.9$, $p = .3$, $R^2 = 36.2\%$). Type of control group did not significantly moderate the treatment effect ($F(2,1) = 0.68$, $p = .65$, $R^2 = 0\%$).

Descriptive analyses revealed a perfect dependence between intervention intensity and type of population ($V = 1$) with all studies with prevention populations evaluating low intensity interventions and all studies with early intervention populations evaluating high intensity interventions. There was a perfect dependence between intervention intensity and type of control group ($V = 1$) with high intensity interventions being only evaluated in comparison to wait-list and placebo controls and low intensity interventions being evaluated in comparison to SC/TAU controls, only. There was a perfect dependence between type of population and type of control group ($V = 1$) with all studies with early intervention populations using wait-list and placebo controls, while all studies with prevention populations used SC/TAU controls.

Follow-up. Out of three studies measuring diagnostic status concerning SSD/FSS at follow-up, appropriate effect size data were available for two studies. Therefore, these data are

synthesized narratively. In the study by Gil-Jardiné et al. (2018), patients at high risk for developing a postconcussion syndrome were treated with a session of either EMDR or reassurance by a therapist in the emergency room. Control subjects received SC/TAU. Diagnostic status was determined via an interview based on the DSM-IV criteria for postconcussion syndrome. Based on a sample of n = 123 and a follow-up length of 3 months, there was a significant effect favoring the intervention groups (RR = 0.54, 95%-CI: [0.37, 0.78]). It is important to note that this effect stems from a worst-case-scenario analysis in which subjects abandoning the intervention protocol due to early discharge or clinical worsening were considered as having an SSD/FSS at follow-up.

Kongsted et al. (2008) examined the effect of oral advice given by a nurse at a home visit to patients presenting with a whiplash injury compared to SC/TAU consisting of an educational pamphlet. These patients were of comparably lower risk for chronic whiplash syndrome since patients at high risk were invited to participate in another trial. Diagnosis was defined via a combination of a neck pain measure and current work status. Based on a sample of 158 subjects and a follow-up length of 12 months, there was no significant effect of the intervention (RR = 1.2, 95%-CI: [0.93, 1.55]).

Although the study by Gatchel et al. (2006) did not provide appropriate effect size data for meta-analytic integration, it reports the effect of the intervention in another effect size metric. Therefore, we describe this study here, too. The study evaluated a combined CBT and biofeedback treatment program for patients suffering from acute jaw pain at high risk for developing a temporomandibular joint disorder. Patients in the control group received no intervention in the context of the trial. Diagnosis was determined by fulfilling the criteria for a pain disorder using the Structured Clinical Interview for DSM-IV. Based on a sample of n = 101 and a follow-up length of 10.5 months, there was a significant positive effect of the intervention (odds ratio = 0.11, 95%-CI: [0.04, 0.29]).

Anxiety.

Post-treatment. Out of four studies measuring anxiety post-treatment, effect size data were available for three of them (n = 237). There was a small and non-significant negative effect ($g = -0.052$, 95%-CI: [-0.33, 0.22], see Figure I11). Heterogeneity was not

significantly different from zero ($Q(2) = 0.34$, $p = .84$) and inconsistency was small to considerable ($I^2 = 0\%$, 95%-CI: [0%, 84.6%]). The resulting 95%-prediction interval ranged from -0.33 to 0.22.

Risk of bias in individual studies. The risk of bias ratings for each domain are depicted in Figure I12.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = -0.018$ (95%-CI: [-3.58, 3.54]). No corrected effect estimate could be computed via 3PSM due to convergence problems.

Additional analyses. No moderator analysis of intervention intensity could be conducted as all studies examined high intensity interventions. No moderator analysis of mean symptom duration could be conducted as no study provided data for this moderator. Type of population did not significantly moderate the treatment effect ($F(1,1) = 68.1$, $p = .077$, $R^2 = 0\%$). No moderator analysis of type of control group could be conducted as all studies examined SC/TAU controls.

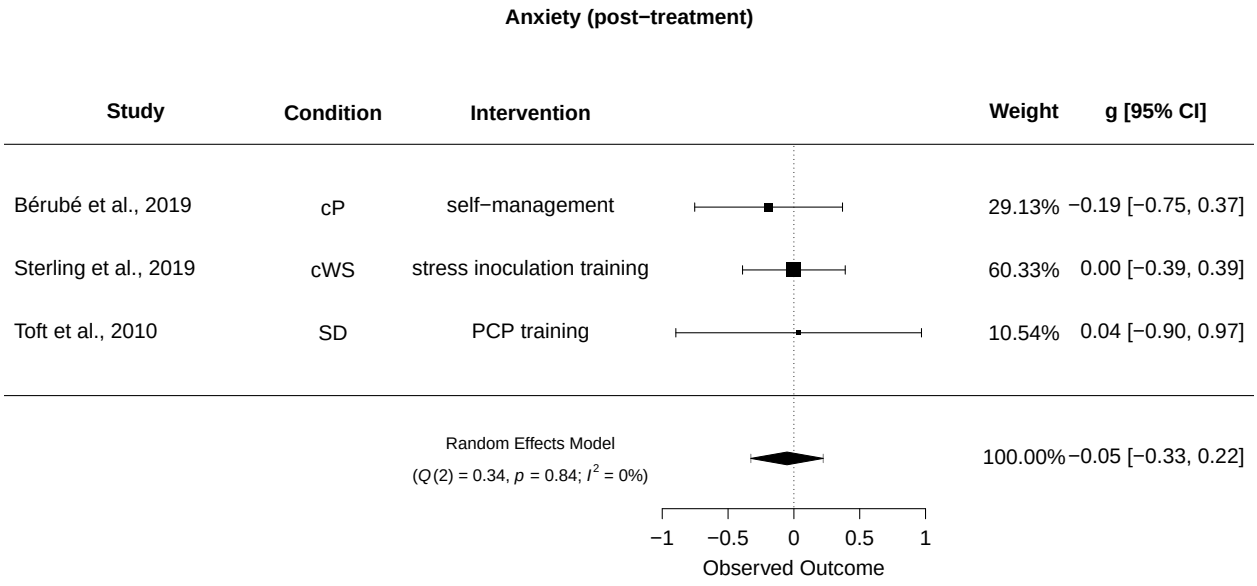

Figure I11. Forest plot of anxiety (post-treatment). $g > 0$ indicates more favorable outcomes in the intervention group. cP: Chronic pain. cWS: Chronic whiplash syndrome. PCP: Primary care physician. SD: Somatoform disorder.

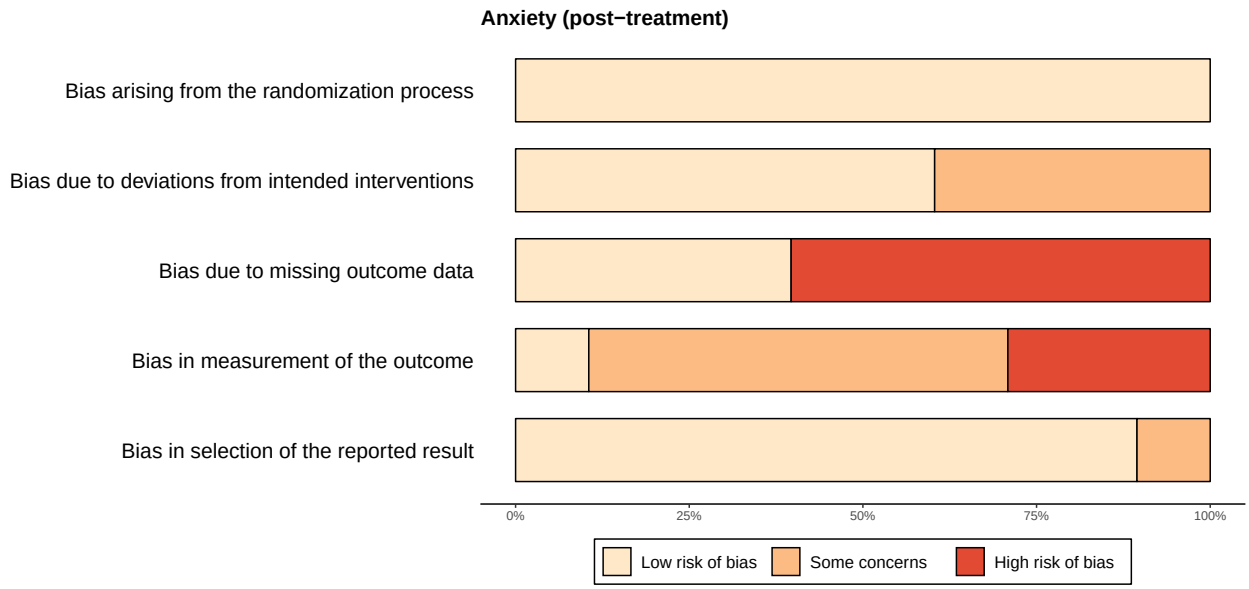

Figure I12. Risk of bias inherent in the summary effect for anxiety (post-treatment). Study-level biases are weighted according to the meta-analytic weights. One cluster-randomized study was included in this meta-analysis (Toft et al., 2010). There was a high risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization in this study (not depicted).

Follow-up. Out of seven studies measuring anxiety at follow-up, effect size data were available for six studies (n = 573). Follow-up length ranged from 1.5 months to 24 months (*Median* = 11.25). There was a small and non-significant negative effect ($g = -0.01$, 95%-CI: [-0.19, 0.17], see Figure I13). Heterogeneity was not significantly different from zero ($Q(5)$ = 3.06, $p = .69$) and inconsistency was small to substantial ($I^2 = 0\%$, 95%-CI: [0%, 65.1%]). The resulting 95%-prediction interval ranged from -0.19 to 0.17.

Risk of bias in individual studies. The risk of bias ratings for each domain are depicted in Figure I14.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.017$ (95%-CI: [-0.58, 0.61]). The 3PSM revealed a corrected effect estimate of $g = 0.002$ (95%-CI: [-0.19, 0.19]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 0.29$, p = .59).

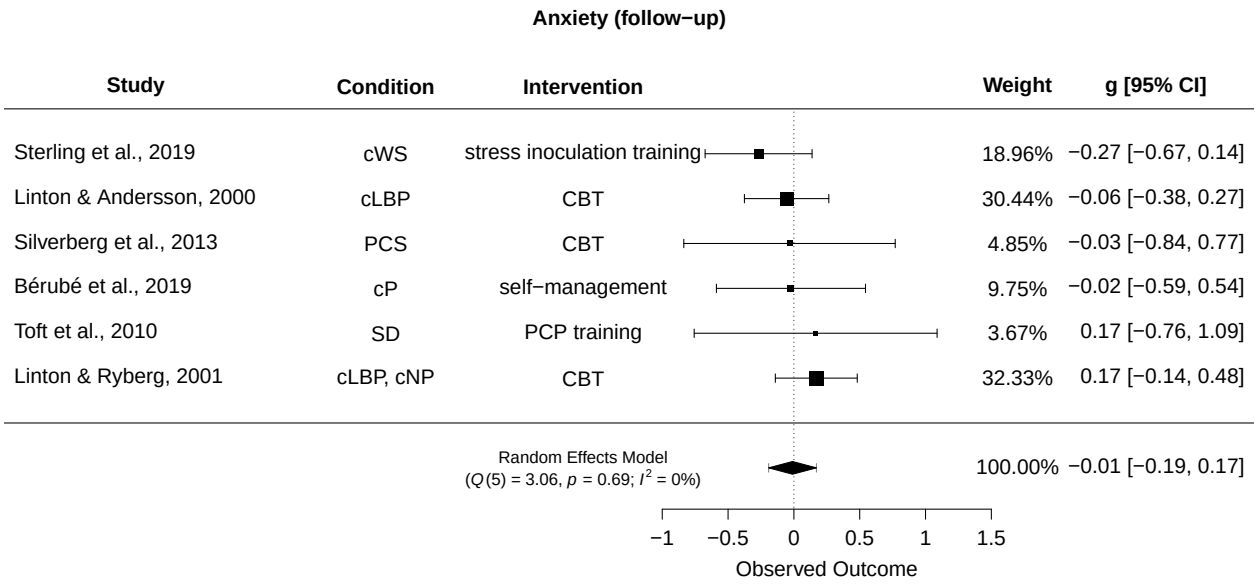

Figure I13. Forest plot of anxiety (follow-up). $g > 0$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. cP: Chronic pain. cWS: Chronic whiplash syndrome. PCP: Primary care physician. PCS: Post-concussion syndrome. SD: Somatoform disorder.

Additional analyses. No moderator analysis of intervention intensity could be conducted as all studies examined high intensity interventions. No moderator analysis of mean symptom duration could be conducted as no study provided data for this variable. Type of population did not significantly moderate the treatment effect ($F(1,4) = 0.007$, $p = .94$, $R^2 = 0\%$). No moderator analysis of type of control group could be conducted as all studies examined SC/TAU controls. Length of follow-up did not significantly moderate the treatment effect ($F(1,4) = 0.29$, $p = .62$, $R^2 = 0\%$).

There was a large positive rank correlation between type of population and length of follow-up ($\rho = .84$).

Depression.

Post-treatment. Out of six studies measuring depression post-treatment, effect size data were available for five studies ($n = 720$). There was a small significant effect ($g = 0.12$, 95%-CI: [0.03, 0.2], see Figure I15). Heterogeneity was not significantly different from zero ($Q(4) = 0.64$, $p = .96$) and inconsistency was small ($I^2 = 0\%$, 95%-CI: [0%, 24%]). The

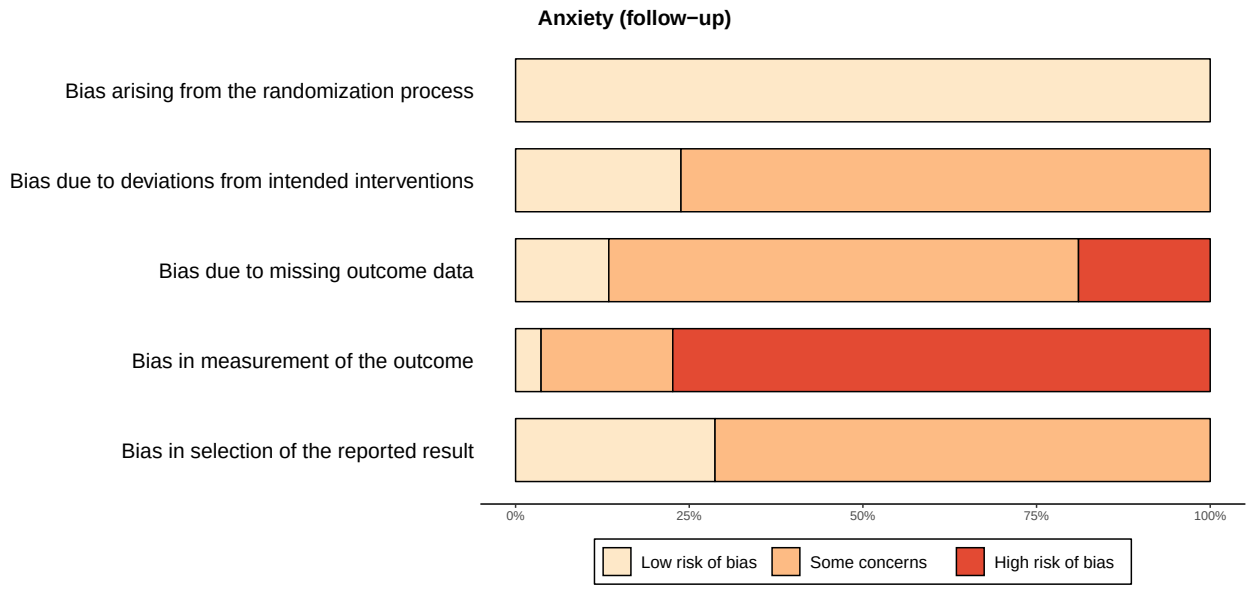

Figure I14. Risk of bias inherent in the summary effect for anxiety (follow-up). Study-level biases are weighted according to the meta-analytic weights. One cluster-randomized study was included in this meta-analysis (Toft et al., 2010). There was a high risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization in this study (not depicted).

resulting 95%-prediction interval ranged from 0.03 to 0.2.

Risk of bias in individual studies. For a summary of risk of bias ratings, see Figure I16.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.12$ (95%-CI: [-0.4, 0.64]). The 3PSM revealed a corrected effect estimate of $g = 0.17$ (95%-CI: [0.046, 0.29]). A likelihood-ratio test did not reveal a significantly better fit of the bias-adjusted model to the data ($\chi^2(1) = 1.32$, $p = .25$).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,3) = 1.34$, $p = .33$, $R^2 = 0\%$). No moderator analysis of mean symptom duration could be computed as there were too few observations ($k = 2$). Type of population did not significantly moderate the treatment effect ($F(1,4) = 0.22$, $p = .67$, $R^2 = 0\%$). Type of control group did not significantly moderate the treatment effect ($F(2,2) = 0.6$, $p = .62$, $R^2 = 0\%$).

There was a small interdependence between intervention intensity and type of population ($V = .25$) with all studies with prevention populations investigating high intensity

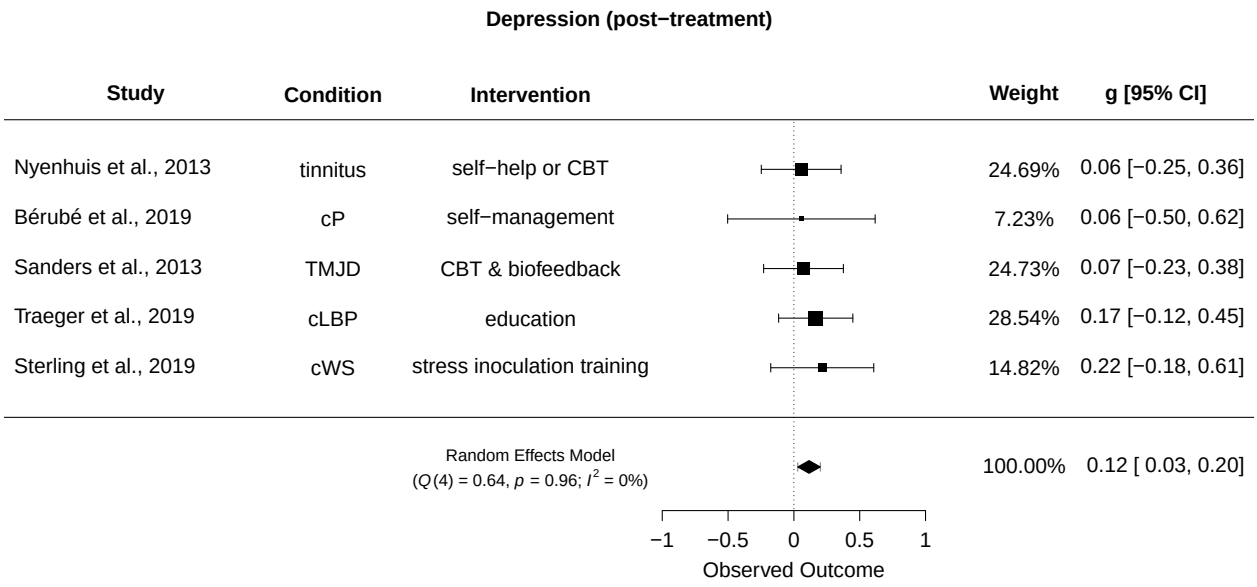

Figure 115. Forest plot of depression (post-treatment). $g > 1$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cP: Chronic pain. cWS: Chronic whiplash syndrome. TMJD: Temporomandibular joint disorder.

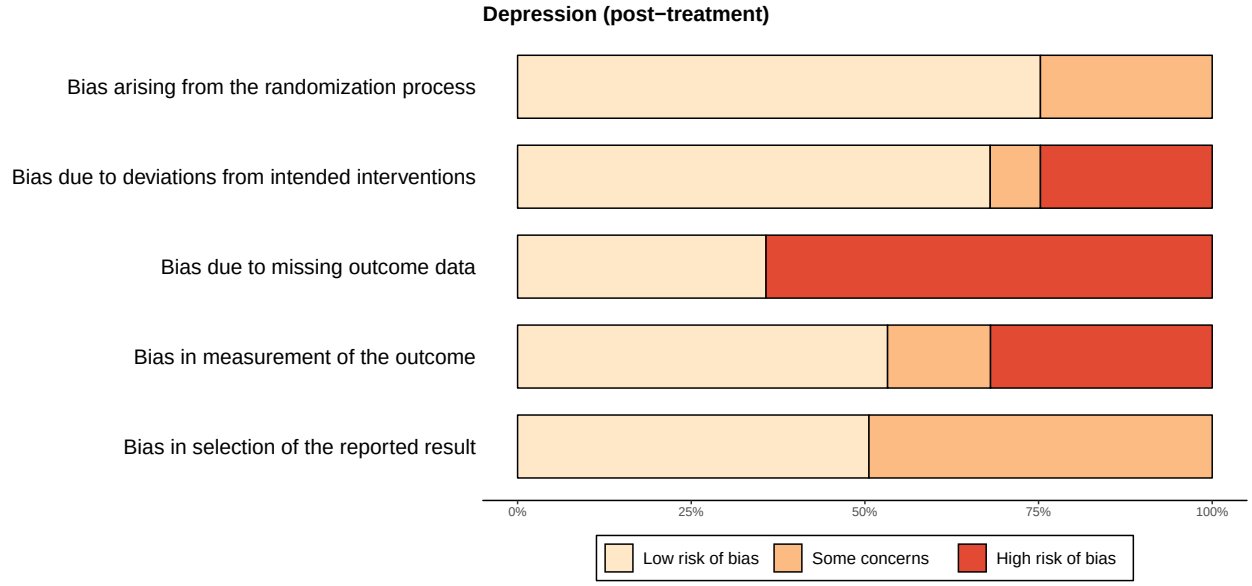

Figure 116. Risk of bias inherent in the summary effect for depression (post-treatment). Study-level biases are weighted according to the meta-analytic weights.

interventions. There was a perfect interdependence between intervention intensity and type of control group ($V = 1$) with high intensity interventions being only investigated in studies with SC/TAU or placebo control groups and low intensity interventions being investigated in studies with no treatment controls, only. There was a large interdependence between type of population and type of control group (V = .61) with all studies with no treatment or placebo comparisons being conducted in early intervention populations.

Follow-up. Out of 10 studies measuring depression at follow-up, effect size data were available for nine studies (n = 1063). Follow-up length ranged from 1.5 months to 12 months (*Median* = 9.5). There was a small and non-significant effect ($g = 0.096$, 95%-CI: [-0.016, 0.21], see Figure I17). Heterogeneity was not significantly different from zero ($Q(8) = 4.83$, $p = .78$) and inconsistency was small to substantial ($I^2 = 0\%$, 95%-CI: [0%, 70.5%]). The resulting 95%-prediction interval ranged from -0.016 to 0.21.

Risk of bias in individual studies. Figure I18 depicts the risk of bias ratings.

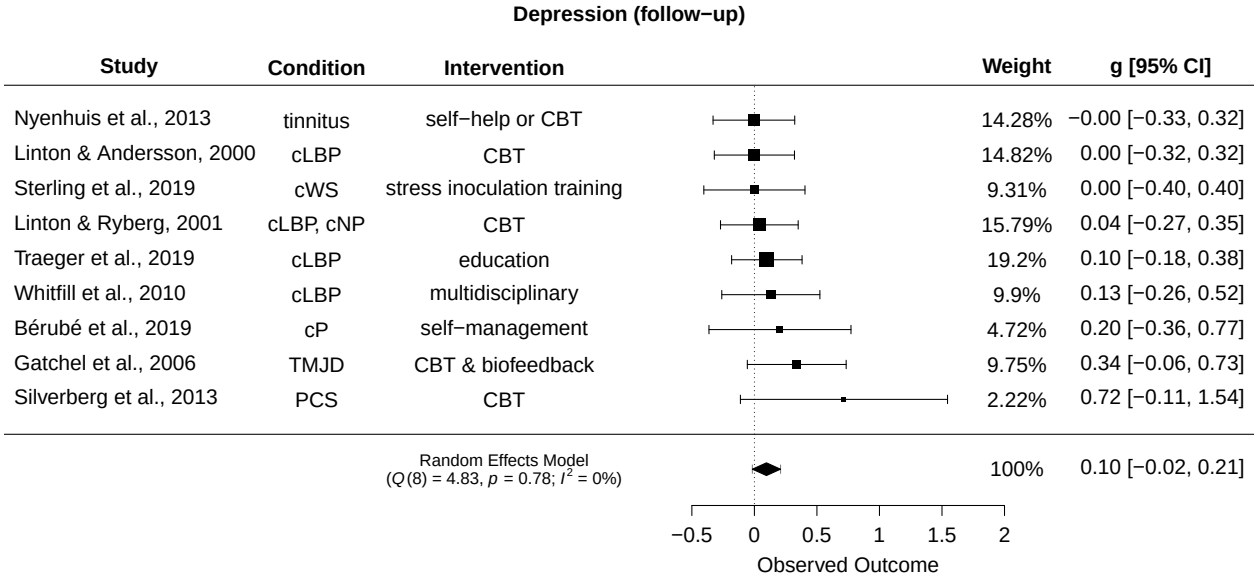

Figure I17. Forest plot of depression (follow-up). $g > 0$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. cP: Chronic pain. cWS: Chronic whiplash syndrome. PCS: Post-concussion syndrome. TMJD: Temporomandibular joint disorder

Meta-bias. The PET-PEESE revealed a corrected effect estimate of g = -0.046 (95%-CI: [-0.19, 0.097]). The 3PSM revealed a corrected effect estimate of $g = 0.14$. A confidence interval could not be computed due to problems with model convergence. A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 1.65$, p = .2).

Additional analyses. Intervention intensity did not significantly moderate the treatment effect ($F(1,7)$ = 0.67, p = .44, R^2 = 100%). Mean symptom duration did not significantly moderate the treatment effect ($F(1,1)$ = 0.037, p = .88, R^2 = 0%). Type of population did not significantly moderate the treatment effect ($F(1,7)$ = 2.83, p = .14, R^2 = 100%). Type of control group did not significantly moderate the treatment effect ($F(2,6)$ = 0.086, p = .92, R^2 = 100%). Length of follow-up did not significantly moderate the treatment effect ($F(1,7)$ = 0.84, p = .39, R^2 = 100%).

There was a medium-sized negative correlation between intervention intensity and mean symptom duration ($r_b = -.49$). There was a small interdependence between intervention intensity and type of population ($V = .19$) with prevention populations being only investigated in studies with high intensity interventions. There was a large interdependence between

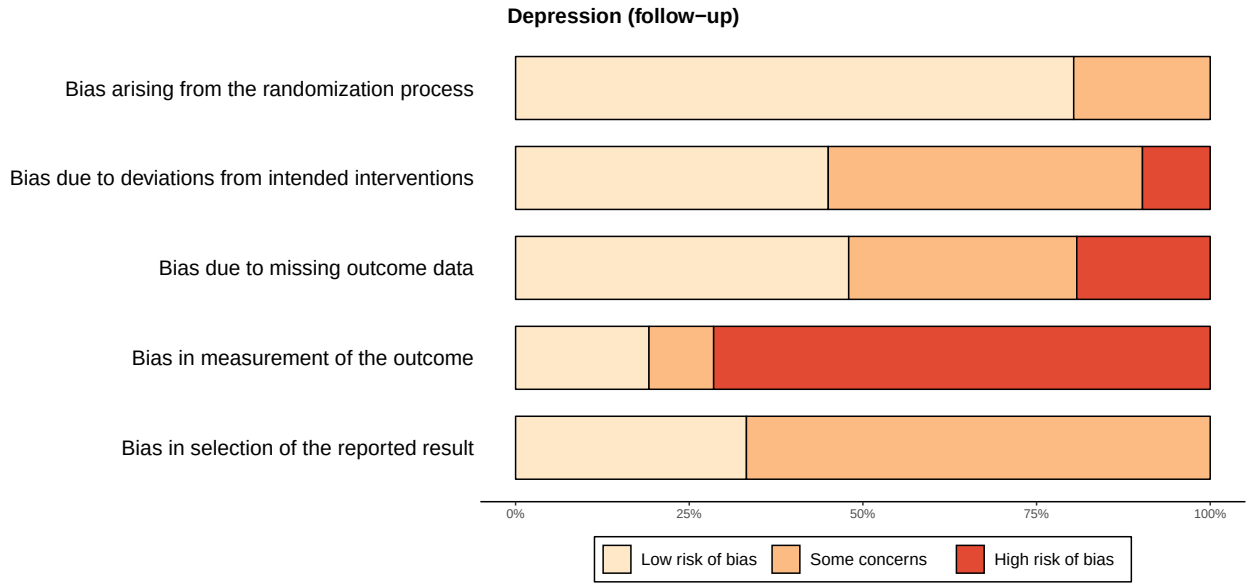

Figure I18. Risk of bias inherent in the summary effect for depression (follow-up). Study-level biases are weighted according to the meta-analytic weights.

intervention intensity and type of control group ($V = .66$) with all studies employing SC/TAU and placebo controls investigating high intensity interventions. There was a small negative correlation between intervention intensity and length of follow-up ($r_b = -.11$). No rank correlation between mean symptom duration and type of population could be computed since all studies providing mean symptom duration data were conducted in early intervention populations. There was a large negative rank correlation between symptom duration and type of comparison ($\rho = -.87$). There was a nearly perfect correlation between mean symptom duration and length of follow-up ($r = .99$). There was a medium-sized interdependence between type of population and type of control group ($V = .38$) with all no treatment and placebo comparisons being conducted in studies investigating early intervention populations. There was a large positive rank correlation between type of population and length of follow-up ($\rho = .58$). There was a small negative rank correlation between type of comparison and length of follow-up ($\rho = -.29$).

Health care utilization.

Post-treatment. Out of four studies measuring health care utilization post-treatment, effect size data were available for none of them.

Follow-up. Out of eight studies measuring health care utilization at follow-up, effect size data were available for three studies (n = 283). Follow-up length ranged from 1.5 months to 12 months (*Median* = 10.5). There was a positive small and significant effect ($g = 0.31$, 95%-CI: [0.18, 0.44], see Figure I19). Heterogeneity was not significantly different from zero ($Q(2)$ = 0.13, $p = .94$) and inconsistency was small to substantial ($I^2 = 0\%$, 95%-CI: [0%, 76.4%]). The resulting 95%-prediction interval ranged from 0.18 to 0.44.

Risk of bias in individual studies. Figure I20 summarizes the risk of bias ratings.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.26$ (95%-CI: [0.15, 0.38]). The 3PSM revealed a corrected effect estimate of $g = 0.82$. A confidence interval could not be computed due to problems with model convergence. A likelihood-ratio test revealed a significantly better fit of the 3PSM to the data ($\chi^2(1) = 5.75$, p = .016).

Additional analyses. No moderator analysis of intervention intensity could be computed as all studies evaluated high intensity interventions. No moderator analysis of mean

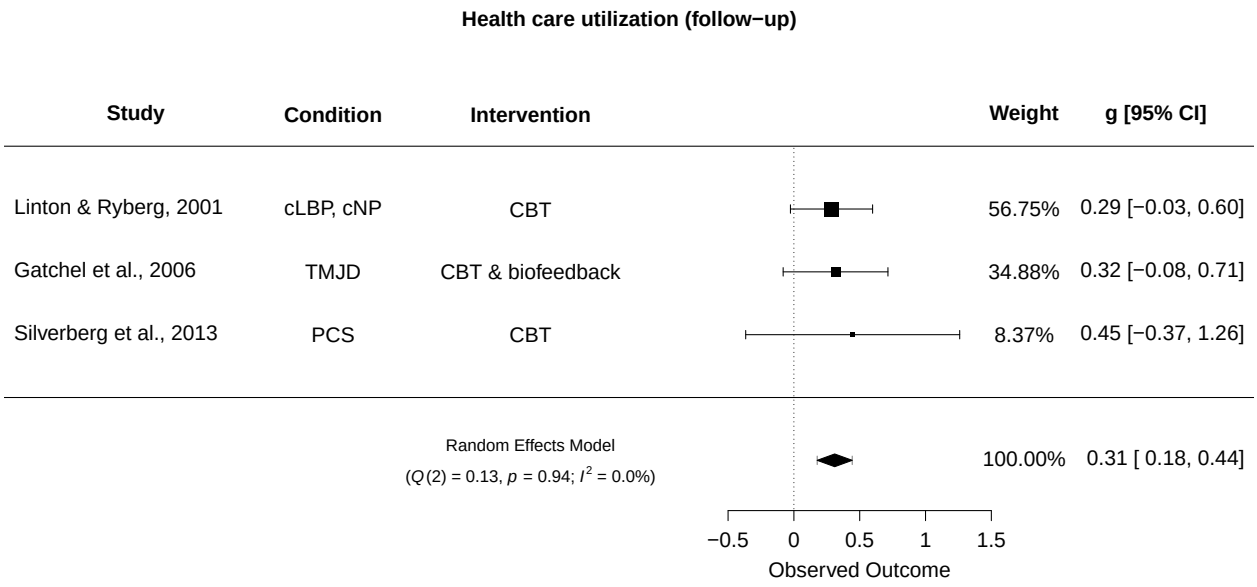

Figure I19. Forest plot of health care utilization (follow-up). $g > 1$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cNP: Chronic neck pain. PCS: Post-concussion syndrome. TMJD: Temporomandibular joint disorder.

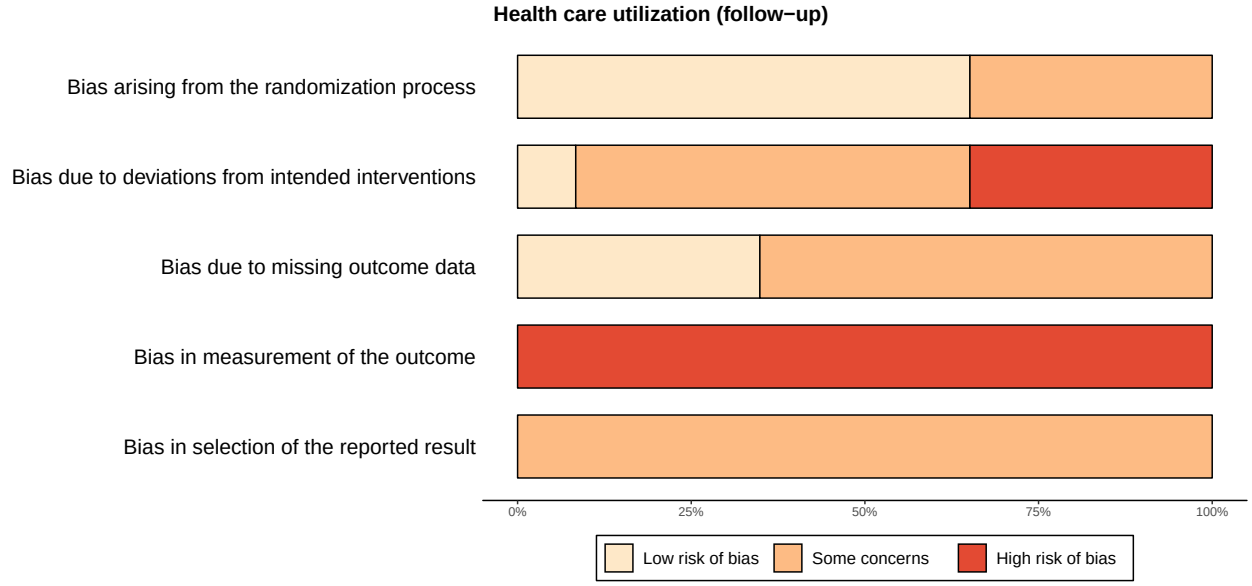

Figure I20. Risk of bias inherent in the summary effect for health care utilization (follow-up). Study-level biases are weighted according to the meta-analytic weights.

symptom duration could be computed as there were too few available studies ($k = 2$). Type of population did not significantly moderate the treatment effect ($F(1,1) = 8.46$, $p = .21$, $R^2 = 0\%$). Type of control group did not significantly moderate the treatment effect ($F(1,1) = 0.012$, $p = .93$, $R^2 = 0\%$). Length of follow-up did not significantly moderate the treatment effect ($F(1,1) = 145.3$, $p = .053$, $R^2 = 0\%$).

There was a large interdependence between type of population and type of control group ($V = .5$) with all no treatment comparisons being conducted in studies with early interventions populations. There was a large positive rank correlation between type of population and length of follow-up ($\rho = .87$). There was a no rank correlation between type of comparison and length of follow-up ($\rho = 0$).

Consumer satisfaction.

Post-treatment. Out of six studies measuring consumer satisfaction post-treatment, appropriate effect size data were available for two studies ($n = 371$). Therefore, the data were synthesized narratively. In the study by Damush et al. (2003b), subjects with acute low back pain participated in a self-management program while control subjects received SC/TAU. Based on a sample of $n = 163$, there was no significant effect of the intervention ($g = -0.02$, 95%-CI: [-0.33, 0.29]).

In the study by Nyenhuis, Zastrutzki, Weise, et al. (2013), subjects suffering from acute tinnitus were treated either with group CBT, bibliotherapy or an online self-help program in the intervention groups. Except for an information sheet concerning the auditory system, tinnitus and treatment options, the control subjects received no treatment. There was a large significant combined effect of the interventions ($g = 1.21$, 95%-CI: [0.89, 1.54], $n = 208$).

Although the other studies did not provide appropriate data for meta-analytic integration, there was other information concerning the consumer satisfaction available. In the study by Gil-Jardiné et al. (2018) evaluating EMDR or reassurance compared to SC/TAU in patients at high risk for post-concussion syndrome, consumer satisfaction was rated on an 11-point numeric rating scale ranging from 0 to 10 with higher values indicating higher satisfaction. There was a median satisfaction of 9.5 (interquartile range (*IQR*): 8 - 10, $n = 34$) in the EMDR group, a median satisfaction of 8.5 (*IQR*: 7.25 - 10, $n = 38$) in the reassurance

group and a median satisfaction of 8 (*IQR*: 6 - 10, *n* = 37) in the control group.

In the study by Karjalainen et al. (2004), consumer satisfaction was rated on the same scale. Subjects were patients suffering from subacute low back pain. The intervention consisted of advice, physiotherapeutic exercises and for a subset of subjects also of a worksite visit by a physiotherapist and a physician. The control group received SC/TAU. Intervention groups resulted in a combined mean satisfaction of 6.15 (range: 0 - 10, *n* = 104), while the SC/TAU group resulted in a mean satisfaction of 4.1 (range: 0 - 10, *n* = 56).

Follow-up. Out of five studies measuring consumer satisfaction at follow-up, effect size data were available for one study. In the study by Damush et al. (2003b, described above), there was no significant difference between the intervention and the control group ($g = 0.098$, 95%-CI: [-0.24, 0.43], *n* = 139) after a follow-up length of 11.25 months.

There were two further studies with relevant data, although they did not report enough data for calculating an effect size. In the study by Karjalainen et al. (2004, described above) there was a combined mean satisfaction of 5.99 (range: 0 - 10, *n* = 103) in the intervention groups and a mean satisfaction of 4.3 (range: 0 - 10, *n* = 53) in the SC/TAU group after at the 24-months follow-up.

In the study by Silverberg et al. (2013), subjects at risk for post-concussion syndrome received six sessions of CBT. Consumer satisfaction was assessed using a 5-point Likert scale. At 1.5 months follow-up, the mean satisfaction in the intervention group was 4.69 (*SD* = 0.48, *n* = 13) indicating high satisfaction. There were no data available for the SC/TAU control group.

Online Supplement J

Sensitivity analyses: exclusion of Janse et al., 2016

Primary outcomes

Somatic symptom severity.

Post-treatment. Out of 18 studies measuring somatic symptom severity post-treatment, effect size data were available for 12 studies (n = 1,931). There was a small and non-significant effect ($g = 0.064$, 95%-CI: [-0.11, 0.24], see Figure J1). Heterogeneity was significantly different from zero ($Q(11) = 29.8$, $p = .002$) and inconsistency was small to considerable ($I^2 = 60.8\%$, 95%-CI: [19.1%, 86.4%]). The resulting 95%-prediction interval ranged from -0.43 to 0.55.

Risk of bias in individual studies. Risk of bias inherent in the summary effect is depicted in Figure J2.

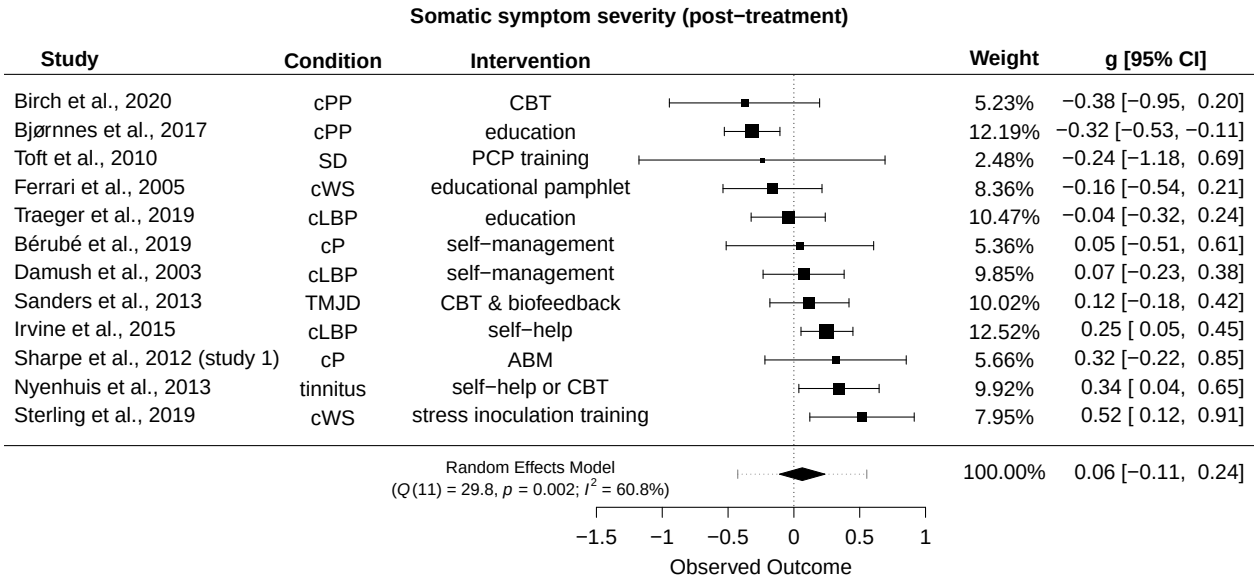

Figure J1. Forest plot of somatic symptom severity (post-treatment). $g > 0$ indicates more favorable outcomes in the intervention group. ABM: Attention bias modification. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cP: Chronic pain. cPP: Chronic postoperative pain. cWS: Chronic whiplash syndrome. PCP: Primary care physician. SD: Somatoform disorder. TMJD: Temporomandibular joint disorder.

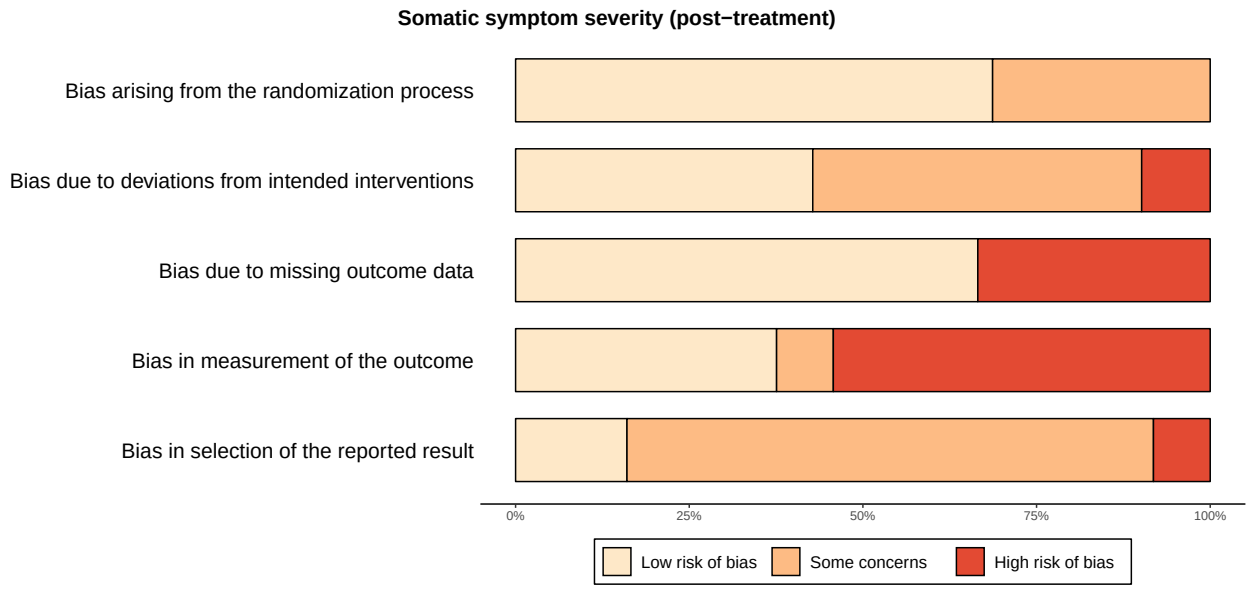

Figure J2. Risk of bias inherent in the summary effect for somatic symptom severity (post-treatment). Study-level biases are weighted according to the meta-analytic weights. One cluster-randomized study was included in this meta-analysis (Toft et al., 2010). There was a high risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization in this study (not depicted).

Meta-bias. The PET-PEESE revealed a corrected effect estimate of $g = 0.044$ (95%-CI: [-0.44, 0.53]). The 3PSM revealed a corrected effect estimate of $g = -0.042$ (95%-CI: [-0.2, 0.11]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 2.72$, p = .099).

Health-related quality of life.

Post-treatment. Out of 16 studies measuring health-related quality of life post-treatment, effect size data were available for 10 studies (n = 4,398). There was a small and non-significant effect ($g = 0.095$, 95%-CI: [-0.1, 0.29], see Figure J3). Heterogeneity was not significantly different from zero ($Q(9)$ = 16.6, $p = .055$) and inconsistency was small to considerable ($I^2 = 45.6\%$, 95%-CI: [0%, 88.5%]). The resulting 95%-prediction interval ranged from -0.34 to 0.53.

Risk of bias in individual studies. Risk of bias inherent in the summary effect is depicted in Figure J4.

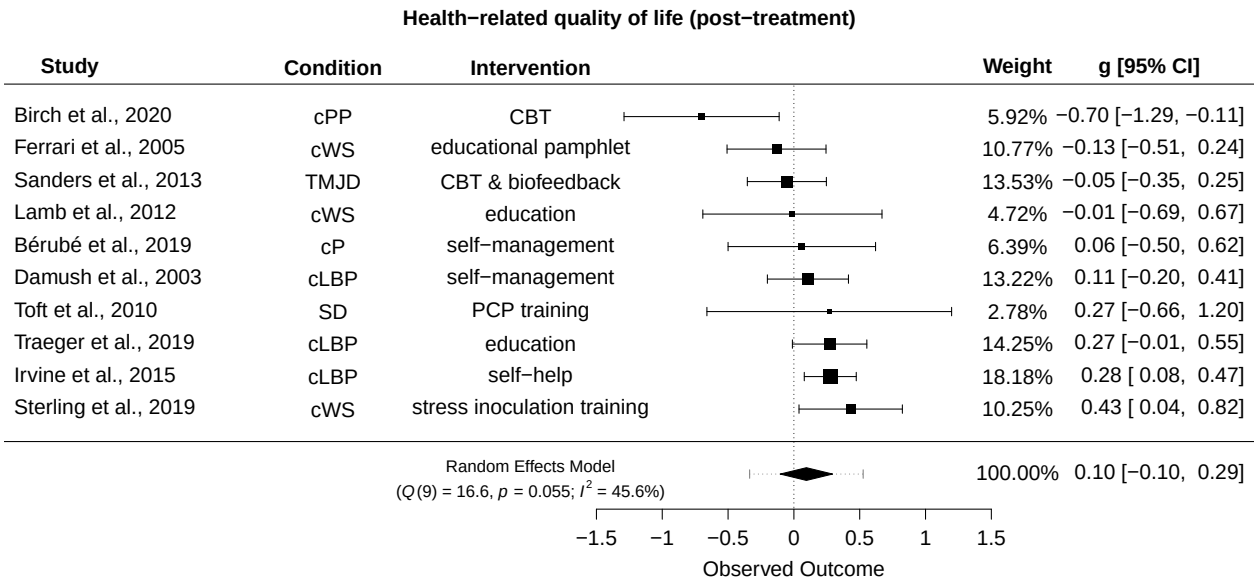

Figure J3. Forest plot of health-related quality of life (post-treatment). $g > 0$ indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cP: Chronic pain. cPP: Chronic postoperative pain. cWS: Chronic whiplash syndrome. PCP: Primary care physician. SD: Somatoform disorder. TMJD: Temporomandibular joint disorder.

Meta-bias. The PET-PEESE revealed a corrected effect estimate of g = 0.21 (95%-CI: [-0.027, 0.45]). The 3PSM revealed a corrected effect estimate of $g = 0.054$ (95%-CI: [-0.13, 0.24]). A likelihood-ratio test did not reveal a significantly better fit of the 3PSM to the data ($\chi^2(1) = 0.81$, p = .37).

Secondary outcomes

Diagnostic status concerning SSD/FSS.

Post-treatment. Three studies measured diagnostic status concerning SSD/FSS post-treatment. Effect size data were available for all of them (n = 327). A random-effects meta-analysis revealed a risk ratio of 1.0003 (95%-CI: [0.5, 1.99], see Figure J5). Heterogeneity was not significantly different from zero ($Q(2)$ = 4.1, $p = .13$) and inconsistency was small to considerable ($I^2 = 51.4\%$, 95%-CI: [0%, 99%]). The resulting 95%-prediction interval ranged from 0.34 to 2.9.

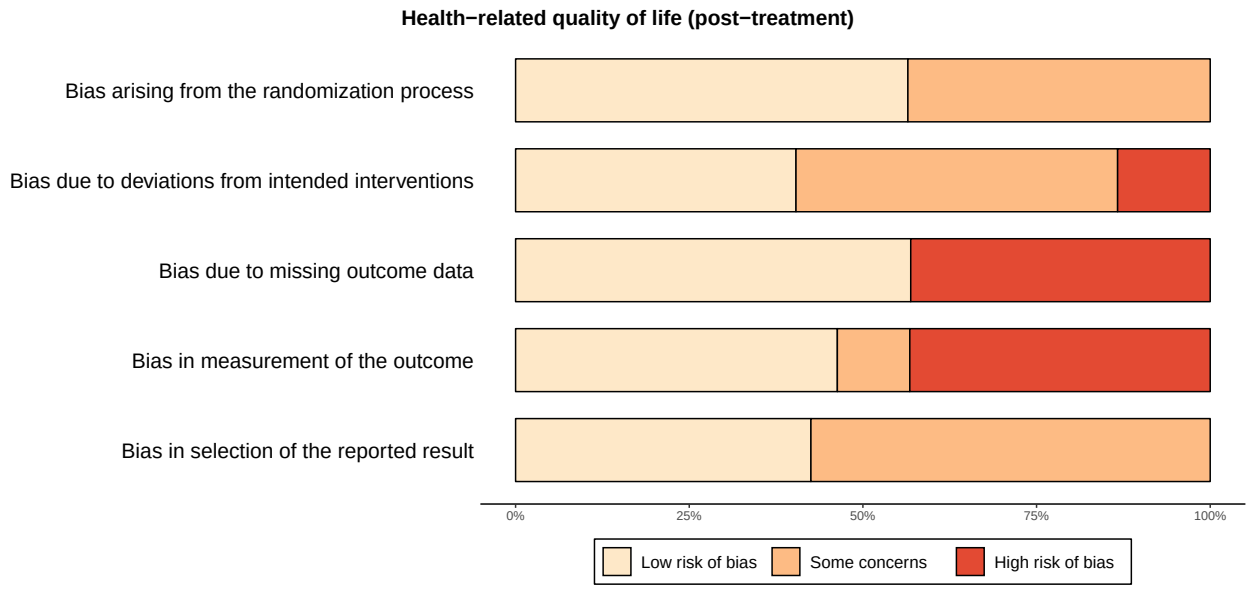

Figure J4. Risk of bias inherent in the summary effect for health-related quality of life (post-treatment). Study-level biases are weighted according to the meta-analytic weights. Two cluster-randomized studies were included in this meta-analysis (Lamb et al., 2012; Toft et al., 2010). While the study by Lamb et al. (2012) was at low risk of bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization, the study by Toft et al. (2010) was at high risk (not depicted).

Risk of bias in individual studies. Figure J6 depicts the risk of bias inherent in the diagnostic status summary effect.

Meta-bias. The PET-PEESE revealed a corrected risk ratio of 2.14 (95%-CI: [0.000001, 3,815,270]). No corrected estimate could be computed using 3PSM due to convergence problems.

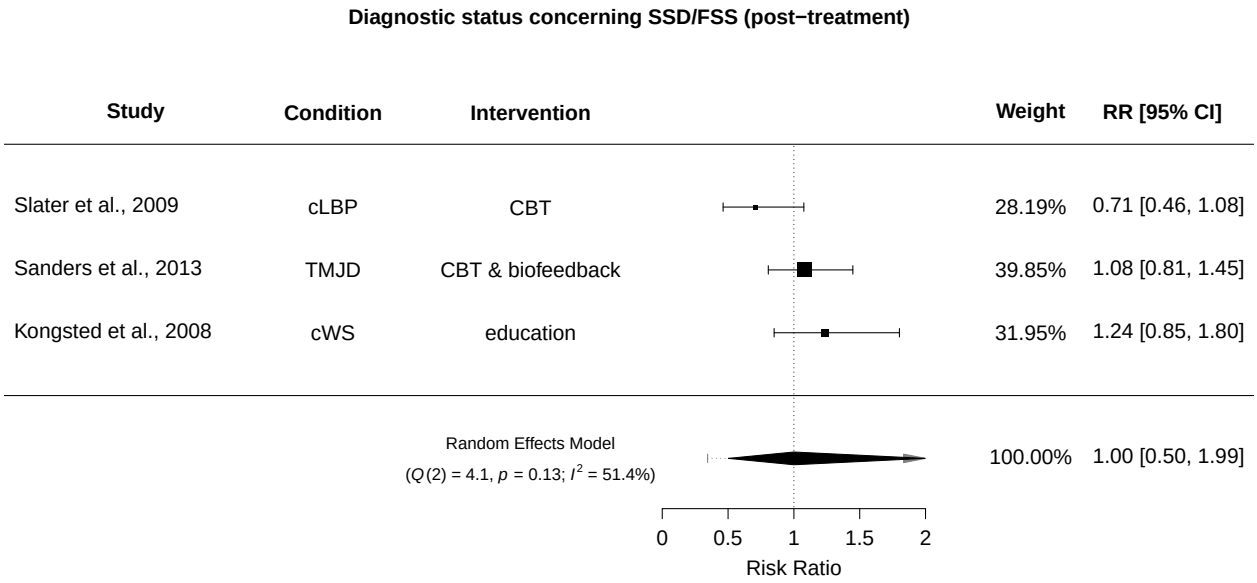

Figure J5. Forest plot of diagnostic status concerning SSD/FSS (post-treatment). *RR* < 1 indicates more favorable outcomes in the intervention group. CBT: Cognitive-behavioral therapy. cLBP: Chronic low back pain. cWS: Chronic whiplash syndrome. TMJD: Temporomandibular joint disorder.

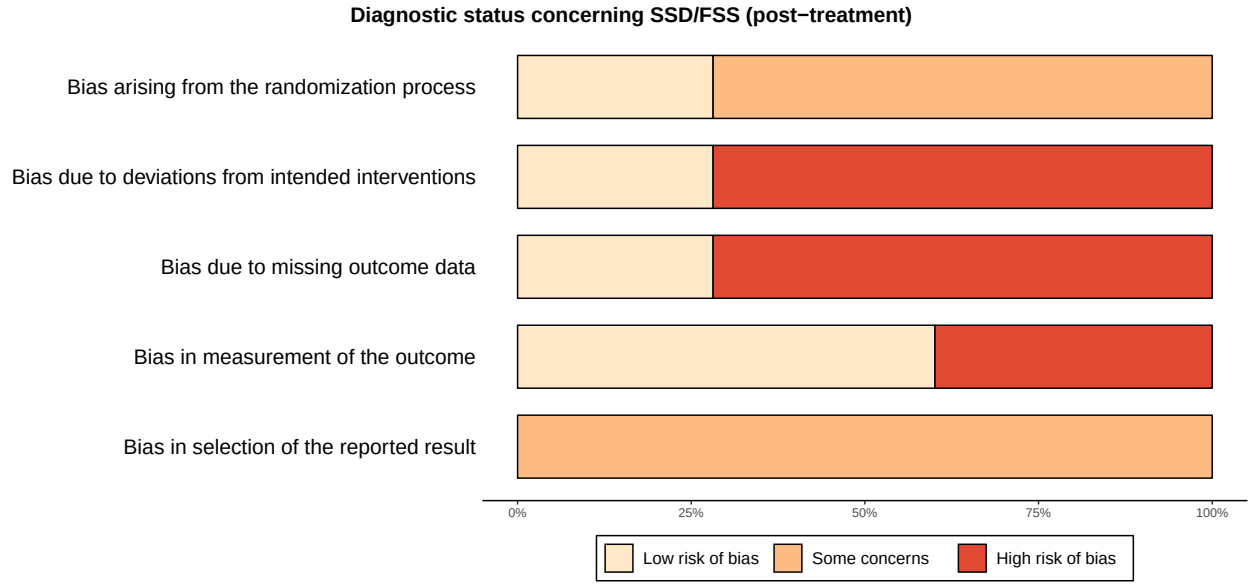

Figure J6. Risk of bias inherent in the summary effect for diagnostic status concerning SSD/FSS (post-treatment). Study-level biases are weighted according to the meta-analytic weights.