PAIN Publish Ahead of Print
DOI: 10.1097/j.pain.0000000000002317

Pragmatic trials of pain therapies: a systematic review of methods.

David Hohenschurz-Schmidt[1], Bethea A. Kleykamp[2], Jerry Draper-Rodi[3], Jan Vollert[1], Jessica Chan[9], McKenzie Ferguson[6], Ewan McNicol[7], Jules Phalip[10], Scott R Evans[8], Dennis C. Turk[11], Robert H Dworkin[2], Andrew S.C. Rice[1]

[1]Department of Surgery & Cancer, Faculty of Medicine, Imperial College London, United Kingdom

[2]University of Rochester School of Medicine and Dentistry, Rochester, NY, USA

[3]University College of Osteopathy, London, United Kingdom

[6]Department of Pharmacy Practice, Southern Illinois University Edwardsville, Edwardsville, Illinois, USA

[7]Department of Pharmacy Practice, MCPHS University, Boston, MA, USA

[8]Biostatistics Center and the Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, George Washington University, Washington, DC, USA

[9]Faculty of Medicine, Imperial College, London, United Kingdom.

[10]European School of Osteopathy, Maidstone, United Kingdom

[11]Department of Anesthesiology & Pain Medicine, University of Washington, Seattle, WA, USA

corresponding author address: Pain Research Group Prof Andrew SC Rice, 4th Floor, Chelsea & Westminster Hospital 369 Fulham Road, London SE10 9NH, United Kingdom, Email: d.hohenschurz-schmidt19@imperial.ac.uk phone: +447708989392, URL: www.imperial.ac.uk/people/d.hohenschurz-schmidt19

Abstract

Pragmatic randomised clinical trials aim to directly inform clinical or health policy decision-making. Here, we systematically review methods and design of pragmatic trials of pain therapies to examine methods, identify common challenges, and areas for improvement.

Seven databases were searched for pragmatic randomised controlled clinical trials which assessed pain treatment in a clinical population of adults reporting pain. All screening steps and data extractions were performed twice. Data were synthesised descriptively and correlation analyses between pre-specified trial features and PRECIS-2 (PRagmatic – Explanatory Continuum Indicator Summary 2) ratings and attrition were performed. Protocol registration: PROSPERO-ID CRD42020178954.

Of 57 included trials, only 21% assessed pharmacological interventions, the remainder physical, surgical, psychological or self-management pain therapies. Three-quarters of the trials were comparative effectiveness designs, often conducted in multiple centres (median: 5; Q1/3: 1, 9.25) and with a median sample size of 234 patients at randomization (Q1/3: 135.5; 363.5). Although most trials recruited chronic pain patients, reporting of pain duration was poor and not well described. Reporting was comprehensive for most general items, while often deficient for specific pragmatic aspects. Average ratings for pragmatism were highest

for treatment adherence flexibility and clinical relevance of outcome measures. They were lowest for patient recruitment methods and extent of follow-up measurements and appointments.

Current practice in pragmatic trials of pain treatments can be improved in areas such as patient recruitment and reporting of methods, analysis and interpretation of data. These improvements will facilitate translatability to other real-world settings – the purpose of pragmatic trials.

Keywords: pain; Clinical trials; Pragmatic Trials; comparative Effectiveness Research; Trial Methodology; systematic review

1. Introduction

Increasingly, alternatives to the classical placebo-controlled randomised clinical trial (RCT) are proposed. The main criticism traditional RCTs concerns the lack of generalisability of research findings due to key aspects of the trial design [98], including exhaustive exclusion criteria (co-morbidity, polypharmacy, psychiatric illness, substance use disorder) [98,107,4,100,121], trial populations differing from the general patient population [72,79], and unrealistic treatment compliance [73,24,60]. Maybe most importantly, what matters to a patient may not have been assessed in an RCT: to be relevant for clinical decision-making, statistical changes in outcome measures need to be reflected in clinically noticeable *and* personally valuable changes in symptoms, quality of life, or disease risk [31,112,25,120]. Even so, the time horizon of a patient's decision is rarely encapsulated by common RCT follow-up periods that are usually six months or less [32]. Despite this lack of

generalizability, RCTs still form the basis of most health policies, medicines regulatory approval and treatment guidelines [89,57,43,45], as they provide a situation in which most factors apart from the intervention are controlled as well as realistically possible.

Pragmatic trial designs have been proposed as a possible remedy to bridge the gap between highly controlled RCTs and clinical practice. The concept of 'pragmatism' refers to the research aim of directly informing a health care or health policy decision, especially in situations where there is a choice between two or more options [17,40,56,103,115]. Importantly, 'pragmatism' in clinical trials is best viewed as a continuum, the two poles being explanatory (efficacy) RCTs and pragmatic (effectiveness) trials [115,68]: many RCTs entail pragmatic elements to increase their external validity, while some 'pragmatic' trials employ methods such as placebo control and blinding [42].

By concept, pragmatic trials are large in scale, embedded into ongoing clinical practice, and frequently investigate complex interventions. Whilst patients are still randomly assigned to treatment groups, they are rarely blinded to their allocation. Also, the treatment protocol is deemed flexible, for example allowing clinicians to adjust drug therapy to individual patients. Outcome measures are thought to reflect what is important in clinical practice, focusing on disability and function, risk-benefit analyses or even cost-effectiveness rather than average pain scores [96,78,17,124,42]. At the extreme end of the explanatory-pragmatic spectrum, pragmatic trials assess outcome data sampled routinely in clinical practice and alter routine care minimally or not at all. To enable clinicians to judge how relevant a study's findings are to a particular clinical scenario, many have called to improve reporting of features associated with the external validity of trials (such as details of the study population, provider expertise, treatment centre volumes, and intervention standardisation)

[11,13,99,33,68,12]. Tools are available to guide pragmatic trial design [114,68] and an extension of the CONsolidated Standards of Reporting Trials statement (CONSORT) exists for pragmatic trials [130].

Aims and objectives

Pragmatic approaches to trial design have been promoted with the goal of increasing the relevance of clinical trials to real world decision-making and policy implementation. To understand the current specifics of trial design, conduct and reporting in the field of pragmatic trials of pain treatments, the objectives of this review were to:

1. Survey the number of randomised controlled trials that are declared to be 'pragmatic' or 'comparative effectiveness' and that are investigating any therapy aimed at pain-reduction in an adult human population experiencing clinical pain.
2. Identify which therapeutic interventions have been assessed in such trials.
3. Evaluate the prevalence of individual design features relating to the concept of 'pragmatism' amongst the included studies.
4. Determine areas for future debate and research within the field of pragmatic trials of pain treatments.

Notably, the aim of this review was not to gauge trials' risk of internal bias or review the effectiveness of treatments.

2. Methods

Protocol registration

A protocol formulated in accordance with the 2015 statement of Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) [76,86] and detailing both the review methods and analysis plan was pre-registered with the International Prospective Register of Systematic Reviews (PROSPERO) prior to commencing data extraction [10] ([55], registration ID: CRD42020178954). Ethical approval was not required.

Eligibility criteria

We reviewed any randomised clinical trial [126], declared by the study authors to be 'pragmatic', 'practical', or 'comparative effectiveness research'. To be included, studies had to investigate people with clinical, i.e., non-experimental pain (including procedure-related pain, irrespective of their age, gender, underlying pathology or the severity and duration of their pain). All interventions aimed at reducing pain in a clinical population or at affecting an outcome measure relevant to the treatment or management of people in pain were eligible, irrespective of treatment setting or delivery format. Trials were included where pain or pain-related measures formed part of the primary analysis, or where the primary aim was to assess endpoints directly relevant to the treatment and management of patients in pain and not administrative processes or diagnosis. No geographic restrictions were applied. Included trials had to have a control or comparison group but the type of comparator was irrelevant for study selection. Within-patient controls were not eligible. Retrospective and observational studies were excluded, as were studies drawing exclusively from registry data. We excluded

feasibility or pilot studies in order to capture the challenges of conducting full-scale pragmatic trials and a minimum of 40 participants per study arm was required. Primary outcome reports had to be published in peer-reviewed sources between January 2018 and March 2020. This timeframe was chosen for several reasons: the rapidly evolving nature of the field [128], the aim to capture the *status quo* in order to inform future methods development, and also because the last milestone paper for the design of pragmatic trials, the PRECIS-2 tool [68], was published in 2015 and we deemed three years to be the minimum amount of time for this recommendation framework to be reflected by the published reports of pragmatic trials. Studies published in the languages English, German, Spanish, Italian, French, and Mandarin were eligible, and others if translations could be obtained. Studies were excluded if no full text could be retrieved, neither online nor through the corresponding author.

Information sources

The following databases were searched from 01 January 2018 to 01 March 2020: MEDLINE, Embase, and PsychInfo (through Ovid interface); the Cochrane Central Register of Controlled Trials (CENTRAL), NIH Clinicaltrials.gov, CINAHL (nursing and allied health, via EBSCO), and the Physiotherapy evidence database (pedro.org.au). As pragmatic trials were expected to be relatively large and costly, we did not anticipate publications in grey literature and no such sources were searched. Reference lists of included studies were reviewed for additional eligible studies. Systematic reviews or meta-analyses were used as sources of further primary studies. We consulted trial registries or contacted authors electronically to identify the trial status when protocols were retrieved. Similarly, authors

were contacted if full reports of potentially eligible trials could not be obtained. For any included study, protocols were consulted for additional information during data extraction.

Search

MeSH or equivalent and text word terms were used and is provided in full as supplement (available at http://links.lww.com/PAIN/B374): pain OR painful conditions (i.e., specific disease names) AND (pragmatic trials OR practical trials OR comparative effectiveness). Limit: human studies, 2018 to current. The search strategy was developed in an iterative manner and under consultation of published literature, designated experts who are part of the research team (pain researchers, trial designers, and therapists), as well as experts in systematic review methodology and database searching. The full search string is provided as supplementary material (available at http://links.lww.com/PAIN/B374).

Study selection

Prior to screening, search results were imported into EndNote (X9) and duplicates removed. For subsequent screening, the studies were exported from Endnote into Covidence, an online platform for systematic reviews (Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org), and another automated de-duplication was performed. Eligibility screening was performed in duplicate, i.e., screened twice by independent reviewers (DHS, AK, EM, JDR, IB, JC, JP). Disagreements were resolved through discussion or, if not possible, by a third party (DHS or RD). In a first step, screening was performed based on study title and abstract. For studies conforming with the eligibility criteria at this

stage and not meeting any of the exclusion criteria, full text publications were accessed and again screened in duplicate.

Data collection process

Like the screening process, the data extraction required a minimum of two independent reviewers. Discrepancies were resolved through discussion and involved a third party if necessary. If available, trial protocols were examined for methods not reported in the trial reports. Missing data which could not be retrieved in this way were recorded as not reported. Where reports were judged to be ambiguous, authors were contacted for clarification. The data extraction form was created iteratively and piloted prior to data extraction.

Data items

The domains of data extraction were source details, funding, trial methods, outcome measures, analysis methods, discussion and contextualisation of information, and reporting. The full extraction table is available as supplementary file (available at http://links.lww.com/PAIN/B374). The extraction of study methods had a triple focus. First, key aspects of pragmatic trial design, conduct, and analysis were extracted. This included eligibility criteria, treatment provision and statistical methods. Secondly, to assess how trialists handled the tension between external and internal trial validity, methods deemed to affect internal trial validity, such as randomization procedures, allocation concealment, and blinding of participants and personnel were extracted [110]. By extracting information on placebo control groups, blinding, and number of trial settings, currently debated areas of

pragmatic trials design were addressed [28,129]. Further, potential shortcomings in randomization were assessed by means of heterogeneity testing between trial groups, performed on baseline age data [21]. Thirdly, to examine how researchers dealt with the specific challenges of pragmatic trials, information on the discussion and methodological treatment of potential heterogeneity between study arms, differences between multiple study centres, differences in therapist expertise, setting resources, treatment flexibility and fidelity, lack of blinding, prolonged follow-up periods, differential attrition, and study cost (i.e., funding information) was extracted. Given that many included studies were expected to be comparative effectiveness trials, it was extracted whether these were designed as superiority or as non-inferiority (equivalence) trials [41].

Complementary to the above-mentioned methods, descriptive data relating to the nine core domains of pragmatic trial design were sampled, as defined by the PRECIS-2 tool [68]. As part of the data extraction process, trials were rated for each of these domains. The PRECIS instrument has been used both during the design phase of trials [58], including in pain research [65], and to retrospectively rate RCTs on a pragmatic – explanatory spectrum [62,14,48,125,92,70,105]. In the latter application, however, some authors have commented on difficulties with inter-rater reliability and missing or unreported data [28,125,80,129]. For this reason, the rating of PRECIS-2 domains was trialed extensively within the team drawing on seminal publications and their explanations [114,62,68,42,128,131], published annotations and examples (https://www.precis-2.org/Trials) as well as the experience of other researchers performing reviews with this tool [105]. Rating occurred in duplicate and inter-rater reliability was assessed. Disagreements of more than one (of five) points were resolved through discussion and / or expert consultation. Otherwise, the average rating was used. Where domains were not applicable or information was insufficient to perform ratings,

domains were left blank, reflecting the current state of the debate in this field [28,129]. PRECIS-2 ratings require comparing a given trial intervention with 'usual care'. In studies where 'usual care' was not described in detail, reviewers had to draw on their own knowledge of the current practice standard. For ambiguous cases, it had been planned to consult national guidelines or clinicians to inform reviewers' conceptions of respective 'usual care' but this was not deemed necessary. Additional data extractions were conducted based on discussions with the review's steering group, including information regarding the content of treatment-as-usual, and details on concomitant pain treatments, risk-benefit, and cost-effectiveness analyses.

Apart from methodological features and PRECIS-2 ratings, recommended reporting items for pragmatic trials were identified, as proposed by the CONSORT statement extension for pragmatic trials [130].

Risk of bias in individual studies

Effect sizes of clinical outcome measures were not extracted nor were potential causes for heterogeneity formally examined (apart from heterogeneity arising from randomization, see below). As the purpose of this review was not to judge clinical or comparative effectiveness, a formal risk of bias assessment was not performed.

Data synthesis

This report was formulated in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [76]. Some subheadings had to be adapted to the purpose of this systematic review of trial methods.

The main results of this review are qualitative and presented using descriptive statistics (means, standard deviations, and percentages of total sample) and appropriate graphs. Additionally, it was assessed if certain trial methods were more prevalent under certain circumstances, using appropriate statistical correlation methods and following a pre-specified analysis plan [55]. Analysing the data in the above way may help inform the design of future trials by highlighting areas for potential conflict and opportunity in trial design.

Risk of bias across studies

Risk of bias across studies was addressed by sampling data on reporting quality, which mainly affected the readers' ability to judge the generalizability of results. By extracting and analysing average and dispersion measures of study participants' age, a heterogeneity meta-analysis between groups sought to identify potential shortcomings in randomization procedures [21,52].

Additional Analyses

Sensitivity analyses examined the above correlations without pre-identified covariates. The only deviation from the analysis plan [55] was the addition of two subgroup analyses, investigating whether PRECIS-2 ratings differed between trials of pharmacological

and non-pharmacological pain therapies as well as between trials of acute versus chronic pain.

Results

Study selection

The search resulted in 769 records after duplicate removal. After excluding 527 records based on titles and abstracts and a further 185 based on assessing the full text for eligibility, 57 individual trials were included in the final sample (Figure 1). Meta-analysis in the sense of descriptive statistics and several correlation analyses was performed on the entire sample of included studies.

[Insert figure 1 about here]

Study characteristics: Descriptive statistics

An overview table of included studies is provided as supplementary material (available at http://links.lww.com/PAIN/B374), specifying each trial's patient population, experimental and comparison interventions, primary outcome measures, and timepoints of primary and longest follow-up. Authors of nine trials described their study as 'comparative effectiveness' but not 'pragmatic', but these trials did not differ significantly from declared pragmatic trials in terms of sample size ($t(55) = 0.13$, $p = 0.9$) nor overall PRECIS-2 score (see below).

Pharmacological treatments for pain were the most studied index treatment (21%), followed by cognitive-behavioural and other psychotherapy approaches (16%), surgery

(12%), acupuncture or acupressure (11%), manual therapies (also 11%), physiotherapy (7%), and others (Table 1). All trials investigated programmes of complex interventions, such as rehabilitation, manual therapy, cognitive-behavioural therapies, various forms of patient management, surgery or drug regimens, or treatment programmes of several modalities.

Concomitant pain treatments were disallowed in six trials (11% of applicable cases, N = 55), either by means of eligibility criteria for participants or after enrolment, and two further trials discouraged patients from seeking treatment outside the trial. However, ten trials did not report whether or not concomitant therapies were permissible.

Where allowed, concomitant treatments were unrestricted in 30 trials (68% of applicable cases) and six trials permitted any concomitant treatment other than those akin or similar to the study interventions. Some trials excluded individual unrelated interventions, such as injections and surgery [88], injections only [87], physiotherapy [16], tricyclic antidepressants [5], or pain medication (not further specified) [59,102]. Only one drug trial made specific allowances for medications and dosages, including some of the same class as study interventions and opioids [35]. Another trial changed the regimen from allowing NSAIDs only in the first six weeks to applying no restrictions thereafter [19]. Thirty trials (68% of applicable cases) reported detail on concomitant pain-reducing treatment actually received, four of which, however, only partly or for specific treatments deemed to relate to the trial intervention (e.g., physiotherapy in a treatment as usual control group of a physiotherapy trial) but not for others (e.g., pain medications [122]).

Whilst pain associated with the musculoskeletal system was most commonly studied (over 60% of trials), diffuse chronic pain conditions such as fibromyalgia were only studied by a single trial [118]. (Table 1).

In more than half the included trials, the patient population consisted of chronic pain patients (Table 1). Whilst this is typically defined as pain lasting for at least three months, patients in most trials had been experiencing pain for several years (supplementary table, available at http://links.lww.com/PAIN/B374). In 12% of trials pain was studied in an acute context, such as pain after injury or associated with medical interventions. Reporting of the sample's exact or even approximate pain duration was, however, poor, with 28% of studies not providing any indication.

[Insert table 1 about here]

The median sample size at the point of randomization was 234 (Q1/3: 135.5; 363.5) with the largest trial featuring a total of 1702 participants [18] and the smallest trial 80 [71].

The median number of trial centres or settings was 5 (Q1/3: 1; 9.25) with one open-label comparative effectiveness trial of drugs for gout flare-ups taking place across 100 general practice clinics across England [95] and 12 (21%) single-centre studies. Seven studies did not report the number of participating treatment centres.

Whilst 37 studies (65%) did not report the number of providers *in the treatment group*, the median number of reported therapy providers was 11 (Q1/3: 6.25; 35.5), with Adams et al. [2] having 400 general practitioners take part in their trial.

Only seven trials (12%) were fully industry-funded [26,29,35,94,97,104,127], 44 trials (77%) had public funders, and five (9%) were funded by mixed sources [50,59,67,83,85]. Funding sources were not reported in one trial [91]. Most trials were conducted in the US (35%), followed by the UK (14%), Australia (7%), and Norway (7%). Four studies (7%) were conducted in East Asia and one in South America [38]. Only one

study was conducted across multiple countries [94]. Protocols were registered for all but one trial [91]. For the purpose of this systematic review, accessing protocols for additional data extraction was deemed necessary in 34 cases (60%).

General trial methods

All but one of the included trials were parallel-group RCTs, with Berdal and colleagues employing a stepped-wedge design [8] and another trial including a cross-over option after 12 weeks [97]. The number and nature of groups differed between trials. Most trials (45; 79%) employed a two-group design, ten trials (18%) had three groups, and two trials (4%) had five groups [118,111] (supplementary table, available at http://links.lww.com/PAIN/B374). More than three-quarters of the trials in the sample were comparative effectiveness (CER) trials, comparing multiple specific interventions or using treatment-as-usual as the comparator. Placebo control groups were employed in 9% trials and 7% a no-treatment control group. One trial each used one of the following alternative comparators: Waitlist controls, advice only, wait & see, and no-treatment without informing patients of a trial being performed (Table 2). Reporting of the content of treatment-as-usual controls is illustrated in table 3.

Participants were reported to be blinded in 13 trials (24% of the trials reporting on participant blinding, n = 54) [2,5,7,8,18,23,29,38,44,82,119,123,127], providers in four (7%, n = 55) [2,5,44,127], and outcome assessors in 45 trials (90%, n = 50). Only five studies reported unblinded assessment (Table 3).

[Insert Table 2 about here]

54 of 57 trials (95%) reported that patients gave informed consent, with none of the trials stating clearly what this information and consent process entailed. In two trials [30,106] consent for patients was waived. In a third trial [2], the unit of randomization were physicians for whom consent was not required; patients, however, did provide informed consent. Lastly, a single article [104] did not report whether or not patients provided informed consent.

Outcome analysis and interpretation

Trials were designed as superiority trials in 52 instances (91%), four (7%) were non-inferiority or equivalence trials [38,63,71,74], and a single study [29] did not report whether or not the trial was designed to show a difference between groups. 'Unsuccessful' superiority trials cannot claim equivalence between interventions [47]; Nonetheless, equivalence or comparative effectiveness was the reported in 9 out of 24 (38%) superiority trials where no significant difference between groups had been demonstrated [6,7,23,50,59,101,104,111,118]. Despite the recommendation to include a third (placebo) control group in non-inferiority trials in order to account for the trial-specific possibility of no demonstrable effect beyond placebo in the control group [41], none of the four non-inferiority trials included a placebo control group.

Out of fifteen trials with multiple outcome measures defined as primary outcomes, nine (60%) did not address the issue of multiplicity in their analysis.

Adherence to reporting guidelines

Adherence to relevant reporting guidelines is presented in table 3.

[Insert Table 3 about here]

Pragmatic trial methods

Average PRECIS-2 ratings

The PRECIS-2 instrument has nine domains that are each rated from 1, indicating a very explanatory design, to 5, indicating a pragmatic approach to a design feature [68].

Inter-rater reliability was moderate for overall PRECIS-2 ratings (ICC 0.73; 95% CI 0.68-0.78, $p < 0.001$), having calculated the intraclass correlation coefficient (ICC) using a two-way mixed-effects model for absolute agreement [90,51,61]. Disagreements had to be resolved through discussion in 22.4% of all instances as initial disagreements exceeded one point on the PRECIS-2 scale. As per protocol, disagreements of a single point were averaged automatically. Assessing inter-rater reliability for individual PRECIS-2 domains, we found moderate (ICC of 0.5 - 0.75) or good (0.75 - 0.9) agreement for all domains but domain 1 (participant eligibility), for which initial agreement was poor (ICC < 0.5) (see Supplement table, available at http://links.lww.com/PAIN/B374) [90].

In our sample of studies, the average rating across all domains was 3.8 (SD 0.62). Only 19 out of 513 overall items were deemed impossible to rate due to required information not being reported (3.7%).

There was no significant difference in overall PRECIS-2 ratings between declared pragmatic trials and those which authors described as ‘comparative effectiveness’ trials and not ‘pragmatic’ ($t(55) = 1.72$, $p = .092$). The only individual domain where there was a significant difference was ‘organization’ ($t(49) = 2.13$, $p = 0.039$), with trials not declared

pragmatic showing less pragmatic features. T-tests for all other domains had *p*-values of > 0.175.

Ratings for individual domains and factors influencing these ratings are presented in more detail below, rating statistics are illustrated by figure 2 and presented in a supplementary table (available at http://links.lww.com/PAIN/B374).

[Insert figure 2 about here]

Eligibility

Assessment of this domain was possible in all instances (M 3.97; SD 1.1; n = 57), even though in only 68% of studies the reporting of eligibility criteria was "*explicitly framed to show the degree to which they included typical participants and/or, where applicable, typical providers (e.g., nurses), institutions (e.g., hospitals), communities (or localities e.g., towns) and settings of care (e.g., different healthcare financing systems)*" [130].

The main reasons for low ratings in this domain was the exclusion of patients with comorbidities common for the specific trial population, which was the case in more than a quarter of all trials (15 studies; 26%). Similarly, common medications were a reason for non-eligibility in seven studies (12%). Once enrolled, patients were advised not to seek additional care outside the trial in 9 studies (16%) [19,23,35,49,88,93,102,111,127]; The report was unclear on that point in 19 instances.

Eligibility criteria for providers, specifically a minimum number of years in practice, could be confirmed in six trials (11%). This information was not reported in 41 studies (72%)

and not relevant for assessment in a further three trials (e.g., where the treatment was automated). A minimum amount of experience with the trial intervention (other than trial-specific training) was required in at least six trials (11%).

Entry criteria for trial centres existed in at least eight cases (14%), not in four cases (7%), and were not reported on in 42 trials (74%). Three further studies took place entirely in the patients' home or another community setting and were thus not relevant for this assessment.

Recruitment

In terms of patient recruitment, convenience sampling was deemed most in line with the principles of pragmatic trials [68]. Instead, however, almost half the trials (25 cases, 44%) resorted to targeted recruitment methods, such as patient identification through records or targeted adverts. A mixed approach was employed by 14%, and recruitment methods were not reported on in another 16% of studies. This reliance on more laborious recruitment strategies is reflected by the fact that PRECIS-2 ratings were lowest on average for this domain (M 3.03; SD 1.6; n = 47); it also points to patient recruitment as a major challenge in pragmatic trials, especially when sought to be performed 'pragmatically', i.e., in tune with every day-practice. In further support of this argument, out of 56 studies which reported a target recruitment number, 15 (27%) did not reach their aim.

Setting

The PRECIS-2 domain 'setting' asks how different the setting of the trial and the usual care settings are. Reporting in line with the CONSORT extension item 21 would help readers to assess generalisability of results: *"Key aspects of the setting which determined the trial results"* [130] were only reported in 21 studies (37%); And a discussion of *"possible differences in other settings where clinical traditions, health service organisation, staffing, or resources may vary from those of the trial"* [130] happened in 19 reports (34%). Whilst explicit considerations of generalisability were often lacking, extrapolation allowed reviewers to rate this PRECIS-2 domain in all but one case, averaging at 3.79 points (SD 1.4; n = 56).

Organization

The PRECIS-2 domain 'organization' compares the provider expertise, resources and the organisation of care delivery in the intervention arm of the trial to those available in usual care [68]. Information on these trial features may help readers to assess the generalisability of trial results to another setting. Nonetheless, 37 trial reports did not indicate the level of experience of those delivering the intervention (69%; n = 54; not relevant in three automated intervention trials). Twenty-nine reports (51%; n = 57) did not state whether or not resources were altered compared to usual care settings in order to implement the intervention. Where this item could be assessed (n = 28), half the studies (50%) did alter resources for the purpose of the trial and half did not. Trial-specific training in an intervention constitutes such an alteration of resources and was part of at least 23 trials (40%). PRECIS-2 ratings for this

domain were lower than for most other domains, specifically a mean of 3.5 (SD 1.3; n = 51). Poor reporting meant that this domain could not be rated in six cases.

Flexibility (Delivery & adherence)

Aspects of the intervention delivery were standardised in 35 trials (61%), with 11 reports not indicating (19%). Out of those trials in which the delivery was reported to be standardised, about one third reported monitoring of the fidelity with which treatments were provided, for example by record checking or video taping of treatment sessions (n = 11, 31%). Where applicable, these features contributed to low PRECIS-2 ratings for this domain and the sample's average was 3.51 (SD 1.2; n = 56) for flexibility in treatment delivery.

For the patients' adherence to treatments and interventions, however, an average of 4.34 points was obtained (SD 1.0; n = 56), meaning that patients were flexible in how they had to follow intervention plans, adherence was encouraged little more than what would be expected in usual practice, and non-adherence rarely meant exclusion from the trial analysis. In fact, post-randomization exclusion criteria such as minimum compliance, absence of adverse events, or other outlier criteria, were only present in a small percentage of trials (n = 5; 9%).

Follow-up

The primary time point at which outcomes were measured was a median of 26 weeks post-randomization (Q1/3: 8; 52 weeks; n = 36) ranging from a single day [104] to five years [6]. However, there were many studies which assessed outcomes over a period of time

[5,9,30,34,97,102,106], at flexible time points [50,81,108], studies which defined several time points as primary [75,111], and several which did not specify a *primary* point of follow-up [3,8,26,59,66,83,109,122]. A better indicator for how long self-declared pragmatic trials are, may thus be the longest time point of follow-up, for which the median was about one year after randomization (Median 50 weeks, Q1/3; 23.25; 52 weeks; n = 56; not reported in one case [50]). The shortest trial assessed peak chest pain during a stenting procedure for acute myocardial infarction, i.e., lasted for no longer than a few hours after the intervention [108]. On the other extreme, Beard et al. [6] are comparing the clinical and cost-effectiveness of total versus partial knee replacements for up to ten years after the event (10-year results not yet published).

Over such potentially long follow-up periods, attrition of study participants can be expected. We found that intervention groups lost an average of 14.9% (SD 12.9, n = 57) of participants until the point of primary follow-up and control groups lost 14.8% (SD 12.78), ranging from no attrition at all to a trial of invasive uterine fibroid surgery where 63% of participants in the intervention group did not complete the trial as per protocol [67].

Across our sample of included studies, there was a significant difference in attrition between groups ($t(56) = 7.16$, $p < 0.001$) and a third of studies reported differential attrition (19; 33%, n = 57). Where there was differential attrition, it was almost as often into the direction of the control group (9 cases, [5,16,81,84,97,101,117,127]) as it was into the direction of the intervention group (11 cases, [7,9,20,67,81,82,85,87,94,111,118,123]), with intervention groups losing an average of 14.9% (SD 12.9) and control groups losing an average of 14.8% (SD 12.8) of participants until the point of primary follow-up ($t(112) =$

0.032, $p = 0.97$), possibly accounting for the fact that both groups were of active interventions in most cases.

Where patients were lost, reasons for drop-out were reported in 35 papers (65%, n = 54).

The PRECIS-2 ratings for the domain 'follow-up' is concerned less with the length of the follow-up period or differential attrition, but rather the frequency and duration of follow-up appointments as well as the intensity of clinical assessments compared to usual care [68]. Based on this, the average rating was 3.24 points (SD 1.3; n = 57), meaning that follow-up was often more elaborate than what would be expected from normal practice.

Outcomes

The choice of outcomes in pragmatic trials should reflect what 'matters' to the patient, choosing direct symptom reports or function-related measures over lab tests, surrogate markers, expert assessment or other external judgement [68]. In our sample of trials, subjective pain ratings, certain condition-related questionnaires, or pain-related functional assessments were the obvious choice [31]. Indeed, on average, trials had the highest rating for this domain (M 4.46; SD 1.0; n = 57). 49 trials obtained a primary outcome through such patient report (86%; n = 57), five trials used a lab or other remote physiological assessment such as radiographs (9%) [35,38,87,91,111], three used a physical or personal assessment (5%) [16,91,101], and one trial each (2%) used data obtained from health records [106] or an objective incident in the medical management of patients, namely the reoccurrence of a medical intervention [67]. As secondary outcomes, objective measures were far more common, being reported in 30 trials (54%) and including measures of

healthcare utilization, physical tests, lab markers, and absence from work. Only half the reports, however, complied with the CONSORT item to justify the chosen outcome and length of follow-up (29 cases, 51%) (Table 3, item 6). Whether or not significant harms or unintended effects occurred was reported in 47 studies (82%) (Table 3, item 19). Harms did occur in 22 of those studies, and in nine of those trials there was a significant difference between groups [23,63,71,81,84,94,95,97,127].

Not affecting PRECIS-2 ratings, but arguably relevant for clinical decision-making are outcome measures and analyses that directly juxtapose treatment risks and benefits [37,36]. None of the included studies employed such composite metrics. Risk-benefit considerations were, however, implicit in 3 trials [64,6,111] assessing high-risk interventions or comparing a high-risk vs. a low risk intervention (e.g., opioid and non-opioid medications). These trials provided extensive data on adverse events. In most other trials, the studied interventions held very little apparent risk to the patients' safety, arguably making risk-benefit analyses less pertinent.

Cost-effectiveness analyses were performed as part of 12 studies (21%) and considered in another eight (14%; either declared in protocol but not reported or considered as part of trial rationale). Downstream healthcare utilization was reported in two trials (4%), allowing for some economic considerations. Again, in some instances one of the tested interventions was so apparently less costly that cost-benefit analyses did not appear warranted if comparative effectiveness or superiority had been shown [74].

Primary analysis

The highest PRECIS-2 rating for this domain is obtained by trials which perform a true intention-to-treat (ITT) analysis for their primary outcome assessment [68], meaning that

all patients randomized are analysed as if treated, irrespective of actual treatment compliance or a failure to attend follow-up assessments (essentially resulting in missing data). 'Pragmatism' in the primary analysis was high, averaging at 4.3 points (SD 1.3, n = 57). The distinction between a true intention-to-treat analysis and a modified ITT [1,53,77] is made clear by the following data: Whilst 48 studies reported to have performed an ITT as primary analysis (84%; not discernible in one instance [18]), 10 of those (21%) excluded participants who did not provide follow up data or had missing data.

Multi-centre trials

At least 45 trials (79%) of our sample were multi-centre studies (with seven studies not reporting the number of participating treatment centres). Despite the possibility of differences between trial centres, for example in terms of case load, resources, and attending patient population, only two studies [49,87] reported having assessed such differences between centres. Those authors also discussed how such differences may have affected the trial results, thus contextualising their findings and enabling the reader to better judge generalisability. Randomization was stratified by site in 15 trials, another method to account for potential differences between study centres.

At least 21% of trials were single-centre trials, which have been highlighted as potentially unpragmatic in recent debates due to supposed low generalizability [28].

Second-level analyses

Baseline heterogeneity

When testing for differences in mean age between intervention and control groups as an indicator of baseline heterogeneity by means of a paired samples t-test, no significant difference was detected ($t(52) = 1.79$, $p = .079$), suggesting that randomization was not systematically biased in this sample of studies.

Preliminary analyses

Testing for potential confounders

There were significant correlations between the overall trial size (total sample size at randomization) and a number of other variables of interest: These included overall PRECIS-2 scores, driven by highly significant correlations with the domains 'setting', 'organization', and 'analysis'. Larger trials were also less likely to show differential attrition, irrespective of in which group most drop-outs occurred (treatment vs. control group) (Table 4). Sample size was thus be included as a covariate of no interest in subsequent analyses of PRECIS-scores and attrition.

[Insert Table 4 about here]

Correlation analyses

A range of planned correlation analyses were performed to identify potential associations amongst different trial features and with ratings of pragmatism, randomization method, and blinding status of participants (Table 5).

Specifically, we asked if overall PRECIS-2 scores were associated with the number of trial centres, the source of trial funding (public, industry, mixed), the primary therapy investigated, the pain condition of participants (index pain disorder), and the employed analysis method (distinguishing a true ITT, modified ITT, and no ITT). Controlling for sample size, only the variable 'Index pain disorder' was associated with PRECIS-2 ratings ($r(54) = -.285$, $p = .033$). Post-hoc analyses to see if specific diagnoses drove this correlation were not possible due to small case numbers in some of the categories.

Whether or not study participants were reported to be blinded to group allocation did not correlate with average PRECIS-2 ratings, the funding source, the size of the trial, or the employed analysis method.

[Insert Table 5 about here]

Attrition

The size of the trial (total n) did not correlate with the percentual attrition, neither in the intervention group nor the control group (r = -0.154 & -0.103, n = 57, p = 0.254 & 0.445, respectively). When ignoring the direction of the attrition, however (i.e., whether more drop-outs occurred in the intervention or the control group), the testing showed that larger trials had less percentual attrition than smaller trials (r = -.360, n = 57, $p = .006$) (Table 4). This latter analysis appears more suitable as the distinction between intervention and control group is somewhat arbitrary in comparative effectiveness trials, where the comparator was often another active intervention.

Subgroup analyses

Given the apparent division of our sample into drug and non-drug trials as well as populations of acute and chronic pain patients, we examined whether total PRECIS-2 ratings differed between these groups of trials. Only trials providing a clear indication of the patient sample's duration of pain and fitting into the categories of acute (< 4 weeks) or chronic (> 3 months) were included into this analysis (n = 38).

One-way analysis of variance revealed no difference in average PRECIS-2 scores between trials of pharmacological and non-pharmacological therapies ($F(1, 55) = 0.27$, $p = 0.6$), with ratings averaging 3.7 (□0.7) and 3.8 (□0.6) out of a maximum score of 5, respectively. Domain-specific ratings only differed for the flexibility with which treatments were delivered, with drug studies allowing significantly less flexibility (2.8 □ 1.2 vs. 3.7 □ 1.6; $F(1,54) = 5.14$, $p = 0.027$).

A significant difference in overall PRECIS-2 scores existed between acute and chronic pain trials ($F(1, 36) = 5.14$, $p = 0.03$), with higher scores in acute trials (4.3 □ 0.7 vs. 3.7 □ 0.6). Exclusion of an outlier, a chronic pain trial with the lowest overall rating [91], did not alter the statistical significance of this result ($F(1,35) = 5.32$, $p = 0.027$). Upon evaluation of individual PRECIS-2 domains, only the domain 'recruitment' differed significantly between groups ($F(1,28) = 5.88$, $p = 0.02$), with chronic pain trials investing more into patient recruitment than acute trials.

3. Discussion

This systematic review of methods describes the current status in the field of declared pragmatic trials in clinical pain therapy research. Such trials typically include several hundred participants, multiple trial centres, and have average follow-up periods of one year. Pragmatic

trials in pain research compare two or more treatments with one another or with 'care as usual'. Treatments are often applied flexibly, and adherence is rarely monitored. Pragmatic trials of pain treatments employ outcome measures that are deemed relevant for clinical decision-making and, in the main, analyse all patients irrespective of treatment compliance or provision of follow-up data. Pragmatic trials in pain research mainly recruit patients living with persistent pain, often musculoskeletal such as back pain or peripheral joint pain, but a small number of pragmatic trials is also conducted in in-patient settings and peri-operatively.

Included trials predominantly investigated complex non-pharmacological interventions rather than drugs. Many manual, rehabilitation, or cognitive-behavioural interventions are already established in routine practice so that equipoise is between two or more alternative (or complementary) treatment options rather than between a new treatment and a placebo. Another driver for pragmatic comparative effectiveness research for non-drug therapies is that these treatments are not subject to drug regulators who require early efficacy and safety signals for market approval. Instead, the non-drug therapy research is produced for clinicians and clinical treatment guidelines, where evidence from comparative effectiveness trials may be acceptable. It is unclear, however, why therapies for centralised pain disorders such as fibromyalgia as well as common complaints such as headaches and neuropathic pain were studied so rarely in pragmatic trials of the last two years.

The present review provides readers with an overview of what is currently called a 'pragmatic trial of pain treatments', enabling them to compare any given trial to this comprehensive description. We did not include a comparison group, for example from a randomly selected sample of pain trials or based on existing reviews. Not only were there feasibility constraints, we also did not want to bias our findings by the selection of

comparison data from non-comparable populations. For example, if we had chosen a systematic review of treatments for neuropathic pain as comparator [39], we would unsurprisingly find large differences to our sample because the neuropathic pain review only studied pharmacological interventions. We are not aware of any reviews of pain treatments that are not restricted to specific populations or interventions.

Apart from describing the 'typical' pragmatic trial in pain research, this systematic review identified several areas for improvement, centred around trial reporting, design, and interpretation.

If the pragmatic aim of 'informing real-world decision-making' [115,129] is to be reached, readers require more information about the environment in which the trial was conducted, including a better description of trial centres, their resources and the typical patient population and diagnoses. This appears particularly important in single-centre trials, making up 21% of our sample, around which there is debate as to whether they can be considered pragmatic at all due to the arguably limited generalisability of results [28,129]. Multi-centre trials, on the other hand, provide the opportunity to assess for differences between study centres and how these factors may have influenced trial results. This was done in two reports only. Additional information about the characteristics of trial centres could facilitate readers' assessment of the applicability of trial results to their particular setting, even when considering single-centre trials. Relatedly, but unlikely specific to pragmatic trials, there is a need to better-describe the population of patients: Too many trials do not indicate the average duration of pain in their sample and many omitted descriptions of the nature or location of pain reported by patients. Similarly, provider characteristics, such as professional qualifications and practical experience with the intervention under investigation, need to be reported. More broadly, trialists cannot assume that readers are aware of the

particularities of the healthcare system or socioeconomic and cultural context in which the trial has been conducted. What constitutes 'care as usual' or how a comparator therapy is implemented may differ widely and is rarely reported in detail. The same is true for concomitant pain treatments, with a fifth of the assessed trials not even indicating whether these were permitted. Detailed information on comparator groups and out-of-study interventions is, however, fundamental to interpreting and understanding the results of any clinical trial and likely more variable in pragmatic trials. Authors are in a unique position to highlight likely similarities with and differences to other potential settings. Another reporting issue is the justification of employed trial methods [80]. For example, why and how did those designing the trial choose certain outcomes and the duration of follow-up periods? Appropriate outcome measures in pain research have been discussed extensively and in an influential publication in 2005 [31]. Possibly, these outcomes have become common practice, making an extensive justification of their choice seem arbitrary. The appropriate length of follow-up periods, on the other hand, is not as well-researched [32]. Authors should thus indicate if the follow-up periods were chosen for clinical reasons, due to patient preferences, or for reasons of trial feasibility, such as funding and drop-out risk.

Our sample of 57 trials obtained an average rating of 3.8 (□0.6) on the PRagmatic - Explanatory Continuum Indicator Summary (PRECIS) 2 instrument [68]. Whilst comparisons with other research fields are difficult, it is noteworthy that this overall score is very similar to ratings of 23 self-declared pragmatic cardiovascular trials which averaged at 3.83 (□0.78) [105].

Domain-specific PRECIS-2 ratings showed that there are a few areas in pain research where 'pragmatic' trial design and conduct are particularly challenging: the relatively large number of patients required often conflicts with the aim to recruit patients in ways comparable to normal practice. Instead of convenience sampling, trialists frequently

implement targeted recruitment strategies such as identification through records and the selective contacting of potentially eligible patients. How much this interfers with the generalisability of trial results remains to be determined. Interestingly, recruitment was more elaborate when chronic pain patients were sampled, demonstrating that challenges and opportunities for pragmatic trials depend on each trial's circumstances and objectives [125]. In the field of pain research, differences appear to exist between trials with acute and chronic patients as well as between drug and non-drug trials.

Relatedly, challenges to and opportunities for the implementation of a pragmatic attitude to trial design can be domain-specific. In general, more pragmatism appears easier to implement in the area of follow-up assessments by means of reducing the frequency and extent of outcome assessments. Nonetheless, follow-up assessments in the present sample often exceeded what would be expected in normal practice, mirroring findings from a small retrospective analysis of weight-loss trials [48]. It could be tempting for trialists to implement more complex and more numerous tests, simply because the opportunity arises. Whilst understandable from a research perspective, extensive outcome testing adds to patient burden and research costs [15,31,40,116]. The extent to which this interfers with patient recruitment and retention is an important question for pragmatic trials and worthy of investigation [27].

Standardisation of treatment delivery was common (61% of the overall sample), and protocol fidelity monitoring occurred in a third of those trials. Ratings for this domain were significantly lower in trials of pharmacological than non-pharmacological treatments, reflecting findings from Koppenaal et al. (2011) who reviewed a set of lifestyle intervention trials and a group of beta-blocker RCTs, few of which, however, declared pragmatic trials [62]. From general practice to complementary and manual therapies, treatments are rarely delivered in an inflexible way. Instead, they are adapted to the patient's needs and preferences, subject to provider expertise and inclinations, as well as influenced by available

resources [22,54,113]. To account for these factors and reconcile them with the need to describe what happened during a trial, instead of artificially restricting the variability in treatment delivery, qualitative research methods may be more appropriate to assess and communicate generalisability. Conversely, such added variability would increase the need for larger samples. Interestingly, treatment adherence was rarely controlled (also compare [70]).

The real or perceived need to control what happens during a trial may also have contributed to the organisation of participating trial centres being more complex and likely more sophisticated than what would be seen in normal practice. Comparably low ratings for this domain were given in a review of RCTs in patients with diabetes [70]. Again, this points to a possible risk to trial design: Should researchers resort to treatment centres and providers who they know can comply with the various requirements of a trial, or do they trust 'normal' practitioners to do the same? Whilst the first option is assumed to further the sucessful recruitment and completion of a trial, it also compromises generalisability, and vice versa for option two. As an encouraging example that large trials can be conducted in non-research facilities, Eklund et al. [34] conducted a trial with 40 chiropracters treating over 300 participants in their private clinics across Sweden.

Another area where the ability of a trial to inform real-world decision-making is potentially hampered by the trial's design is the implementation of placebo control groups and, relatedly, blinding of participants and providers. As Dal-Ré and colleagues [28] point out, these aspects are not part of normal clincial practice and the authors argue that any trial employing them is inherently explanatory. In our review, five trials (9%) employed a placebo control group. Participant blinding was performed in 13 trials, representing a quarter of all trials which reported on participant blinding, and providers were blinded to group allocation in four trials. In the debate on whether these design features preclude labelling a trial 'pragmatic', Zwarenstein et al. [129] respond that, for example in scenarios where

patient or provider subjectivity needs to be excluded as a source of apparent effectiveness, such studies can still inform real-world decision-making, the main intention behind pragmatism in trial design. The pain field with its predominatly subjective outcome measures offers illustrative examples of this reasoning, such as Bayer et al. [5], a self-declared pragmatic trial comparing an off-label beta blocker vs. placebo in the prevention of vestibular migraine, or the CSAW trial of Beard et al. [7], which was the first placebo-controlled trial for subacromial decompression surgery, demonstrating no benefit of real surgery over the surgical placebo (exploratory arthroscopy). Following Dal-Ré's reasoning, however, by employing a third, no-treatment arm and clearly demonstrating a marked placebo effect of both interventions, the CSAW trial had a strong explanatory component that was not reflected in its PRECIS-2 score of 4.33. On the other hand, the results of this trial are clearly relevant to clinical decision-making given that decompression surgery is (still) common practice. It appears therefore that a pragmatic intention is compatible with elements of mechanistic, explanatory studies but that these instances should be clearly highlighted alongside PRECIS-2 ratings to understand the reasoning behind the trial design (also see [82]).

Many of the above consideration point to difficulties when applying the PRECIS-2 instrument to trials design. When understood as an 'incentive' during the planning of a trial, higher ratings in each domain may conflict with internal validity requirements of a trial and the developers rightly point out that high ratings are not an end in themselves [68,131]. Despite being a scale, PRECIS-2 may have contributed to a false dichotomy. Often, trial methods are discussed as either pragmatic or explanatory [28]. Rather than the design, however, it is the trial's objectives that make it pragmatic or explanatory and trial methods simply follow the need to answer pragmatic research questions in a methodologically sound manner [103,42]. Further, when employed retrospectively, the comparison of trials from

different fields may be challenging, with, for example, provider training and fidelity monitoring being much more pertinent issues in complex intervention trials than in pharmacological studies. For the present purpose, however, discrepancies in such ratings allowed for a nuanced discussion of the potential reasons, again highlighting that PRECIS-2 ratings require context.

For future methodological work on pragmatic trials for pain therapies, it is worthwhile to contextualize the inter-rater reliability of our PRECIS-2 ratings. In general, our overall moderate agreement compares favourably to the 2017 PRECIS-2 validation study of Loudon et al. [69] that found good inter-rater reliability for three domains and moderate reliability for the remaining six [90], but we achieved much smaller confidence intervals in our study (Supplementary table 3, available at http://links.lww.com/PAIN/B374). Interestingly, rating a sample of 15 trial protocols from a variety of fields, the test raters in Loudon's study had most difficulty agreeing on ratings for the domains recruitment and intervention adherence, whilst in our study the rating of domain 1 (participant eligibility) was most ambiguous, possibly underlining the need for authors of pragmatic trials to more clearly report if and how their trial population generalizes to the target population of the intervention in routine practice. In our study, domains 2 (recruitment) and 4 (organization) had the most missing data due to insufficient information in protocols and trial reports (48 and 47 complete ratings, respectively, less after reconciliation); for domain 4 (organization) Loudon et al. also had the highest percentage of missing data, attesting to suboptimal reporting of this information in many trials. Our approach of detailed preparation and training of those researchers who performed the PRECIS-2 ratings, plus the averaging of discrepancies of a single point, led to moderate agreements and very feasible reconciliation process where only a fifth of items required a mostly brief discussion and usually without involvement of a third party. However,

the fact that initial inter-rater reliability was nonetheless only moderate, plus the fact that about 4% of domains could not be rated even after discussion, testifies to the inherent challenges of retrospective PRECIS-2 ratings and the need to improve trial reporting to facilitate such assessments in the future [28].

Apart from reporting and design considerations for pragmatic trials, this review raises concerns regarding the analysis and interpretation of trial results. As most pragmatic trials are comparative effectiveness studies and mostly designed to show a difference between group means (superiority trials), authors need to be explicit about the clinical significance of differences, if detected, and cannot claim 'equivalence' if the trial failed to show a significant difference. The latter occurred in over a third of 24 non-significant superiority trials in this sample, much higher than the 10% found in a review of 76 reports of pain therapy trials with non-significant primary analyses [47]. If designed as non-inferiority or equivalence trials, trial designers need to establish assay sensitivity, ideally by including a third, no-treatment or placebo control group [41]. Whilst only four non-inferiority trials were included in the present sample, none of them complied with this recommendation, again making it difficult to interpret the results. Lastly, what authors understand as 'intention-to-treat analysis' (ITT) differs, with 20% of self-declared ITT analyses excluding participants who did not provide follow-up data or where data were missing. The use of such modified ITT analysis and incorrect labelling has direct implications for the interpretation and meta-analysis of results [1,53,77] and mirrors the findings of a 2014 review of phase II and III trials of pain treatments [46].

4. Limitations

We excluded 14 studies because the authors did not employ the terms 'comparative effectiveness', 'pragmatic' or 'practical' in reference to their own study. We pointed out in our protocol that "*This review will only capture trials which have been declared as 'pragmatic', 'practical' or 'comparative effectiveness' by their authors, i.e., publications which contain this or related terms in the title or abstract*." We acknowledged that "*relying on author self-report might result in the inclusion of studies which score low on current tools for the evaluation of pragmatic aspects of trial design (specifically PRECIS-2) as well as the omission of trials not explicitly declared 'pragmatic' but in fact conforming with many criteria of pragmatic trials*". A future sensitivity analysis may wish to examine of including these trials would have affected the results of this systematic review.

Further, the decision to include trials which were declared comparative effectiveness trials but not necessarily declared pragmatic trials may have resulted in the inclusion of trials that were not explicitly designed as pragmatic trials. An area where this may have had an effect is compliance with pragmatic trial reporting guidelines. Also, authors may label their study pragmatic without considering that this should mean a trial with the potential to directly inform clinical decision-making [103,42,129]. Indeed, the lowest-rated study in our sample by Qi et al. may be such a case of conceptual misapplication [91]. Lastly, there likely are declared pragmatic trials that are not randomized and these were outside the scope of this review.

5. Conclusion

In summary, this systematic review provided a comprehensive snapshot of the current practice in the pragmatic design of comparative effectiveness and other pragmatic trials of pain treatments. Such trials typically include several hundred participants, numerous sites, are

publicly funded, and assess complex interventions for the treatment and management of pain, predominantly chronic pain. These trials have long follow-up periods, employ clinically relevant outcome measures, and resemble usual care in the extent to which patients are required to adhere to treatments. The resources employed for patient recruitment and the intensity of follow-up often pre-empted higher ratings of 'pragmatism'. The included trials comply well with basic reporting guidelines but the assessment of generalisability is frequently hampered by poor reporting of design features relevant to pragmatic trials. Overall, the challenges and opportunities for pragmatic trial design are likely largely dependent on an individual trial's objectives and circumstances. There are no recommendations for trial designers regarding how to navigate these challenges, on balancing internal and external validity and on harnessing the potential for pragmatic trials to provide highly clinically relevant insights in a trustworthy manner.

This review ascertained the prevalence of self-declared pragmatic, practical or comparative effectiveness trials for the treatment and management of patients with pain and will inform future development of and guidance on trial methods designed to enhance real-world application of trial findings.

6. Acknowledgements and Disclosures

Thank you to Mrs Emma Shawn for her assistance in devising the search strategy.

Mr. Hohenschurz-Schmidt reports personal fees from the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT), during and for the conduct of the study. Bethea A. Kleykamp has received in the past 36 months support from

the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION) public-private partnership with the United States Food and Drug Administration and compensation for medical writing from Palladian Associates, HayesInc/TractManager, and PBS Next Avenue. In addition, between 2014 and 2018 during her previous employment at the consulting firm, PinneyAssociates, she provided consulting advice to pharmaceutical companies, the e-cigarette company NJOY, and the tobacco company RAI Services Company on non-combustible tobacco products including e-cigarettes; Dr. Ferguson reports other funds from ACTTION, during the conduct of the study; Dr. Vollert reports personal fees from Casquar, personal fees from Vertex Pharmaceuticals, outside the submitted work; Dr. Evans reports personal fees from Takeda / Millennium, personal fees from Pfizer, personal fees from Roche, personal fees from Novartis, personal fees from ACTTION, personal fees from Genentech, personal fees from Amgen, personal fees from American Statistical Association, personal fees from FDA, personal fees from Osaka University, personal fees from Nationa Cerebral and Cardiovascular Center of Japan, personal fees from NIH, personal fees from Society for Clinical Trials, personal fees from Statistical Communications in Infectious Diseases (DeGruyter), personal fees from AstraZeneca, personal fees from Teva, personal fees from Austrian Breast & Colorectal Cancer Study Group (ABCSG)/Breast International Group (BIG) and the Alliance Foundation Trials (AFT) , personal fees from Taylor and Francis, personal fees from Vir, personal fees from Shire, personal fees from Alexion, personal fees from Gilead, personal fees from Clinical Trials Transformation Initiative, personal fees from Tracon, personal fees from Deming Conference, personal fees from Antimicrobial Resistance and Stewardship Conference, personal fees from Advantagene, personal fees from Cardinal Health, personal fees from Microbiotix, personal fees from Stryker, personal fees from Atricure, personal fees from BENEFIT, personal fees from Roivant, personal fees from Neovasc, personal fees from

Nobel Pharma, personal fees from Horizon, personal fees from Roche, personal fees from Rakuten, personal fees from Duke University, personal fees from U. of PENN, personal fees from Takeda , personal fees from Nuvelution, personal fees from Abbvie, personal fees from Clover, personal fees from FHI Clinical, personal fees from Lung Biotech, personal fees from SAB Biopharm, personal fees from CIOMS, personal fees from SVB LEERINK, grants from NIH, outside the submitted work; Dr. Dworkin reports grants from US Food and Drug Administration, during the conduct of the study; personal fees from serving on advisory boards or consulting on clinical trial methods from Abide, Acadia, Adynxx, Analgesic Solutions, Aptinyx, Aquinox, Asahi Kasei, Astellas, AstraZeneca, Biogen, Biohaven, Boston Scientific, Braeburn, Cardialen, Celgene, Centrexion, Chromocell, Clexio, Collegium, Concert, Confo, Decibel, Dong-A, Editas, Eli Lilly, Ethismos (equity), Eupraxia, Glenmark, Grace, Hope, Immune, Lotus, Mainstay, Merck, Neumentum, Neurana, NeuroBo, Novaremed, Novartis, Olatec, Pfizer, Phosphagenics, Quark, Reckitt Benckiser, Regenacy (also equity), Relmada, Sanifit, Scilex, Semnur, SIMR Bio, SK Life Sciences, Sollis, SPRIM, Teva, Theranexus, Trevena, Vertex, and Vizuri, outside the submitted work; Dr. Turk reports personal fees from Pfizer, personal fees from Lilly, personal fees from GSK, outside the submitted work; and is the Editor-in-Chief of the Clinical Journal of Pain. Dr. Rice reports personal fees from IMMPACT, during the conduct of the study; personal fees from Imperial College Consultants, from Spinifex, outside the submitted work; In addition, Dr. Rice has a patent WO 2005/079771 pending, and a patent EP13702262.0/ WO2013 110945 pending; Dr Draper-Rodi, Dr McNicol, Mrs Chan, and Mr Phalip have nothing to disclose.

References

[1] Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. BMJ 2010;340:c2697.

[2] Adams AS, Schmittdiel JA, Altschuler A, Bayliss EA, Neugebauer R, Ma L, Dyer W, Clark J, Cook B, Willyoung D, Jaffe M, Young JD, Kim E, Boggs JM, Prosser LA, Wittenberg E, Callaghan B, Shainline M, Hippler RM, Grant RW. Automated symptom and treatment side effect monitoring for improved quality of life among adults with diabetic peripheral neuropathy in primary care: a pragmatic, cluster, randomized, controlled trial. Diabet Med 2019;36:52–61.

[3] Aragonès E, Rambla C, López-Cortacans G, Tomé-Pires C, Sánchez-Rodríguez E, Caballero A, Miró J. Effectiveness of a collaborative care intervention for managing major depression and chronic musculoskeletal pain in primary care: A cluster-randomised controlled trial. J Affect Disord 2019;252:221–229.

[4] Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet Lond Engl 2012;380:37–43.

[5] Bayer O, Adrion C, Al Tawil A, Mansmann U, Strupp M, PROVEMIG investigators. Results and lessons learnt from a randomized controlled trial: prophylactic treatment of vestibular migraine with metoprolol (PROVEMIG). Trials 2019;20:813.

[6] Beard DJ, Davies LJ, Cook JA, MacLennan G, Price A, Kent S, Hudson J, Carr A, Leal J, Campbell H, Fitzpatrick R, Arden N, Murray D, Campbell MK, TOPKAT Study Group. The clinical and cost-effectiveness of total versus partial knee

replacement in patients with medial compartment osteoarthritis (TOPKAT): 5-year outcomes of a randomised controlled trial. Lancet Lond Engl 2019;394:746–756.

[7] Beard DJ, Rees JL, Cook JA, Rombach I, Cooper C, Merritt N, Shirkey BA, Donovan JL, Gwilym S, Savulescu J, Moser J, Gray A, Jepson M, Tracey I, Judge A, Wartolowska K, Carr AJ, CSAW Study Group. Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. Lancet Lond Engl 2018;391:329–338.

[8] Berdal G, Bø I, Dager TN, Dingsør A, Eppeland SG, Hagfors J, Hamnes B, Mowinckel P, Nielsen M, Sand-Svartrud A-L, Slungaard B, Wigers SH, Hagen KB, Dagfinrud HS, Kjeken I. Structured Goal Planning and Supportive Telephone Follow-up in Rheumatology Care: Results From a Pragmatic, Stepped-Wedge, Cluster-Randomized Trial. Arthritis Care Res 2018;70:1576–1586.

[9] Blödt S, Pach D, Eisenhart-Rothe S von, Lotz F, Roll S, Icke K, Witt CM. Effectiveness of app-based self-acupressure for women with menstrual pain compared to usual care: a randomized pragmatic trial. Am J Obstet Gynecol 2018;218:227.e1-227.e9.

[10] Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, Stewart L. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev 2012;1:2.

[11] Bornhöft G, Maxion-Bergemann S, Wolf U, Kienle GS, Michalsen A, Vollmar HC, Gilbertson S, Matthiessen PF. Checklist for the qualitative evaluation of clinical

studies with particular focus on external validity and model validity. BMC Med Res Methodol 2006;6:56.

[12] Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, for the CONSORT NPT Group. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. Ann Intern Med 2017;167:40.

[13] Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, for the CONSORT Group. Extending the CONSORT Statement to Randomized Trials of Nonpharmacologic Treatment: Explanation and Elaboration. Ann Intern Med 2008;148:295.

[14] Bratton DJ, Nunn AJ. Alternative approaches to tuberculosis treatment evaluation: the role of pragmatic trials. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis 2011;15:440–446.

[15] Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. BMJ 2019;364:k5267.

[16] Cederbom S, Leveille SG, Bergland A. Effects of a behavioral medicine intervention on pain, health, and behavior among community-dwelling older adults: a randomized controlled trial. Clin Interv Aging 2019;14:1207–1220.

[17] Chalkidou K, Tunis S, Whicher D, Fowler R, Zwarenstein M. The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. Clin Trials 2012;9:436–446.

[18] Cherkin D, Balderson B, Wellman R, Hsu C, Sherman KJ, Evers SC, Hawkes R, Cook A, Levine MD, Piekara D, Rock P, Estlin KT, Brewer G, Jensen M, LaPorte A-M,

Yeoman J, Sowden G, Hill JC, Foster NE. Effect of Low Back Pain Risk-Stratification Strategy on Patient Outcomes and Care Processes: the MATCH Randomized Trial in Primary Care. J Gen Intern Med 2018;33:1324–1336.

[19] Chesterton LS, Blagojevic-Bucknall M, Burton C, Dziedzic KS, Davenport G, Jowett SM, Myers HL, Oppong R, Rathod-Mistry T, van der Windt DA, Hay EM, Roddy E. The clinical and cost-effectiveness of corticosteroid injection versus night splints for carpal tunnel syndrome (INSTINCTS trial): an open-label, parallel group, randomised controlled trial. Lancet Lond Engl 2018;392:1423–1433.

[20] Chung VC, Wong CH, Wu IX, Ching JY, Cheung WK, Yip BH, Chan KL, Cheong PK, Wu JC. Electroacupuncture plus on-demand gastrocaine for refractory functional dyspepsia: Pragmatic randomized trial. J Gastroenterol Hepatol 2019;34:2077–2085.

[21] Clark L, Fairhurst C, Hewitt CE, Birks Y, Brabyn S, Cockayne S, Rodgers S, Hicks K, Hodgson R, Littlewood E, Torgerson DJ. A methodological review of recent meta-analyses has found significant heterogeneity in age between randomized groups. J Clin Epidemiol 2014;67:1016–1024.

[22] Cohen DJ, Crabtree BF, Etz RS, Balasubramanian BA, Donahue KE, Leviton LC, Clark EC, Isaacson NF, Stange KC, Green LW. Fidelity Versus Flexibility: Translating Evidence-Based Research into Practice. Am J Prev Med 2008;35:S381–S389.

[23] Cohen SP, Bicket MC, Kurihara C, Griffith SR, Fowler IM, Jacobs MB, Liu R, Anderson White M, Verdun AJ, Hari SB, Fisher RL, Pasquina PF, Vorobeychik Y. Fluoroscopically Guided vs Landmark-Guided Sacroiliac Joint Injections: A Randomized Controlled Study. Mayo Clin Proc 2019;94:628–642.

[24] Coleman EA, Chugh A, Williams MV, Grigsby J, Glasheen JJ, McKenzie M, Min S-J. Understanding and Execution of Discharge Instructions. Am J Med Qual 2013;28:383–391.

[25] Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Qual Life Res 2018;27:33–40.

[26] Côté P, Boyle E, Shearer HM, Stupar M, Jacobs C, Cassidy JD, Carette S, van der Velde G, Wong JJ, Hogg-Johnson S, Ammendolia C, Hayden JA, van Tulder M, Frank JW. Is a government-regulated rehabilitation guideline more effective than general practitioner education or preferred-provider rehabilitation in promoting recovery from acute whiplash-associated disorders? A pragmatic randomised controlled trial. BMJ Open 2019;9:e021283.

[27] Courtright KR, Halpern SD, Joffe S, Ellenberg SS, Karlawish J, Madden V, Gabler NB, Szymanski S, Yadav KN, Dember LM. Willingness to participate in pragmatic dialysis trials: the importance of physician decisional autonomy and consent approach. Trials 2017;18:474.

[28] Dal-Ré R, de Boer A, James SK. The design can limit PRECIS-2 retrospective assessment of the clinical trial explanatory/pragmatic features. J Clin Epidemiol 2020. doi:10.1016/j.jclinepi.2020.03.027.

[29] Darlow B, Stanley J, Dean S, Abbott JH, Garrett S, Wilson R, Mathieson F, Dowell A. The Fear Reduction Exercised Early (FREE) approach to management of low back pain in general practice: A pragmatic cluster-randomised controlled trial. PLoS Med 2019;16:e1002897.

[30] Dissing KB, Hartvigsen J, Wedderkopp N, Hestbæk L. Conservative care with or without manipulative therapy in the management of back and/or neck pain in Danish children aged 9-15: a randomised controlled trial nested in a school-based cohort. BMJ Open 2018;8:e021358.

[31] Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, Kerns RD, Stucki G, Allen RR, Bellamy N, Carr DB, Chandler J, Cowan P, Dionne R, Galer BS, Hertz S, Jadad AR, Kramer LD, Manning DC, Martin S, McCormick CG, McDermott MP, McGrath P, Quessy S, Rappaport BA, Robbins W, Robinson JP, Rothman M, Royal MA, Simon L, Stauffer JW, Stein W, Tollett J, Wernicke J, Witter J. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. PAIN 2005;113:9.

[32] Dworkin RH, Turk DC, Peirce-Sandner S, Baron R, Bellamy N, Burke LB, Chappell A, Chartier K, Cleeland CS, Costello A, Cowan P, Dimitrova R, Ellenberg S, Farrar JT, French JA, Gilron I, Hertz S, Jadad AR, Jay GW, Kalliomäki J, Katz NP, Kerns RD, Manning DC, McDermott MP, McGrath PJ, Narayana A, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Reeve BB, Rhodes T, Sampaio C, Simpson DM, Stauffer JW, Stucki G, Tobias J, White RE, Witter J. Research design considerations for confirmatory chronic pain clinical trials: IMMPACT recommendations. PAIN® 2010;149:177–193.

[33] Dworkin RH, Turk DC, Peirce-Sandner S, Burke LB, Farrar JT, Gilron I, Jensen MP, Katz NP, Raja SN, Rappaport BA, Rowbotham MC, Backonja M-M, Baron R, Bellamy N, Bhagwagar Z, Costello A, Cowan P, Fang WC, Hertz S, Jay GW, Junor R, Kerns RD, Kerwin R, Kopecky EA, Lissin D, Malamut R, Markman JD, McDermott MP, Munera C, Porter L, Rauschkolb C, Rice ASC, Sampaio C,

Skljarevski V, Sommerville K, Stacey BR, Steigerwald I, Tobias J, Trentacosti AM, Wasan AD, Wells GA, Williams J, Witter J, Ziegler D. Considerations for improving assay sensitivity in chronic pain clinical trials: IMMPACT recommendations. PAIN 2012;153:1148–1158.

[34] Eklund A, Jensen I, Lohela-Karlsson M, Hagberg J, Leboeuf-Yde C, Kongsted A, Bodin L, Axén I. The Nordic Maintenance Care program: Effectiveness of chiropractic maintenance care versus symptom-guided treatment for recurrent and persistent low back pain—A pragmatic randomized controlled trial. PLOS ONE 2018;13:e0203029.

[35] Emery P, Horton S, Dumitru RB, Naraghi K, van der Heijde D, Wakefield RJ, Hensor EMA, Buch MH. Pragmatic randomised controlled trial of very early etanercept and MTX versus MTX with delayed etanercept in RA: the VEDERA trial. Ann Rheum Dis 2020;79:464–471.

[36] Evans SR, Bigelow R, Chuang-Stein C, Ellenberg SS, Gallo P, He W, Jiang Q, Rockhold F. Presenting Risks and Benefits: Helping the Data Monitoring Committee Do Its Job. Ann Intern Med 2019;172:119–125.

[37] Evans SR, Follmann D. Using Outcomes to Analyze Patients Rather than Patients to Analyze Outcomes: A Step Toward Pragmatism in Benefit:Risk Evaluation. Stat Biopharm Res 2016;8:386–393.

[38] de-Figueiredo FED-//-L. Apical periodontitis healing and postoperative pain following endodontic treatment with a reciprocating single-file, single-cone approach: A randomized controlled pragmatic clinical trial. PloS One 2020;15:e0227347.

[39] Finnerup NB, Attal N, Haroutounian S, McNicol E, Baron R, Dworkin RH, Gilron I, Haanpää M, Hansson P, Jensen TS, Kamerman PR, Lund K, Moore A, Raja SN, Rice ASC, Rowbotham M, Sena E, Siddall P, Smith BH, Wallace M. Pharmacotherapy for neuropathic pain in adults: a systematic review and meta-analysis. Lancet Neurol 2015;14:162–173.

[40] Fiore LD, Lavori PW. Integrating Randomized Comparative Effectiveness Research with Patient Care. N Engl J Med 2016;374:2152–2158.

[41] Food and Drug Administration. Non-Inferiority Clinical Trials to Establish Effectiveness Guidance for Industry. 2016. Available: https://www.fda.gov/media/78504/download.

[42] Ford I, Norrie J. Pragmatic Trials. N Engl J Med 2016;375:454–463.

[43] Frieden TR. Evidence for Health Decision Making — Beyond Randomized, Controlled Trials. N Engl J Med 2017;377:465–475.

[44] Friedman BW, Cisewski D, Irizarry E, Davitt M, Solorzano C, Nassery A, Pearlman S, White D, Gallagher EJ. A Randomized, Double-Blind, Placebo-Controlled Trial of Naproxen With or Without Orphenadrine or Methocarbamol for Acute Low Back Pain. Ann Emerg Med 2018;71:348-356.e5.

[45] Gastaldon C, Mosler F, Toner S, Tedeschi F, Bird VJ, Barbui C, Priebe S. Are trials of psychological and psychosocial interventions for schizophrenia and psychosis included in the NICE guidelines pragmatic? A systematic review. PloS One 2019;14:e0222891.

[46] Gewandter JS, McDermott MP, McKeown A, Smith SM, Pawlowski JR, Poli JJ, Rothstein D, Williams MR, Bujanover S, Farrar JT, Gilron I, Katz NP, Rowbotham MC, Turk DC, Dworkin RH. Reporting of intention-to-treat analyses in recent analgesic clinical trials: ACTTION systematic review and recommendations. PAIN 2014;155:2714–2719.

[47] Gewandter JS, McKeown A, McDermott MP, Dworkin JD, Smith SM, Gross RA, Hunsinger M, Lin AH, Rappaport BA, Rice ASC, Rowbotham MC, Williams MR, Turk DC, Dworkin RH. Data Interpretation in Analgesic Clinical Trials With Statistically Nonsignificant Primary Analyses: An ACTTION Systematic Review. J Pain 2015;16:3–10.

[48] Glasgow RE, Gaglio B, Bennett G, Jerome GJ, Yeh H-C, Sarwer DB, Appel L, Colditz G, Wadden TA, Wells B. Applying the PRECIS Criteria to Describe Three Effectiveness Trials of Weight Loss in Obese Patients with Comorbid Conditions. Health Serv Res 2012;47:1051–1067.

[49] Goertz CM, Long CR, Vining RD, Pohlman KA, Walter J, Coulter I. Effect of Usual Medical Care Plus Chiropractic Care vs Usual Medical Care Alone on Pain and Disability Among US Service Members With Low Back Pain: A Comparative Effectiveness Clinical Trial. JAMA Netw Open 2018;1:e180105.

[50] Griswold D, Learman K, Kolber MJ, O'Halloran B, Cleland JA. Pragmatically Applied Cervical and Thoracic Nonthrust Manipulation Versus Thrust Manipulation for Patients With Mechanical Neck Pain: A Multicenter Randomized Clinical Trial. J Orthop Sports Phys Ther 2018;48:137–145.

[51] Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol 2012;8:23–34.

[52] Hicks A, Fairhurst C, Torgerson DJ. A simple technique investigating baseline heterogeneity helped to eliminate potential bias in meta-analyses. J Clin Epidemiol 2018;95:55–62.

[53] Higgins JPT, Thomas J, Cumpston M, Li T, Page MJ, Welch VA eds. Cochrane Handbook for Systematic Reviews of Interventions, Version 6.0. 2nd ed. Chichester, UK: John Wiley & Sons, 2019 p. Available: www.training.cochrane.org/handbook. Accessed 2 Sep 2019.

[54] Higgs J, Titchen A. Rethinking the Practice-Knowledge Interface in an Uncertain World: a Model for Practice Development. Br J Occup Ther 2001;64:526–533.

[55] Hohenschurz-Schmidt D, Kleykamp B, Draper-Rodi J, Voller J, Dworkin R, Rice AS. Pragmatic trials of pain therapies: a systematic review of methods (Protocol ID: CRD42020178954). PROSPERO 2020. Available: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020178954.

[56] Institute of Medicine. Initial National Priorities for Comparative Effectiveness Research. Washington, DC.: The National Academies Press, 2009 p. doi:10.17226/12648.

[57] Irving M, Eramudugolla R, Cherbuin N, Anstey KJ. A Critical Review of Grading Systems: Implications for Public Health Policy. Eval Health Prof 2017;40:244–262.

[58] Johnson KE, Neta G, Dember LM, Coronado GD, Suls J, Chambers DA, Rundell S, Smith DH, Liu B, Taplin S, Stoney CM, Farrell MM, Glasgow RE. Use of PRECIS

ratings in the National Institutes of Health (NIH) Health Care Systems Research Collaboratory. Trials 2016;17:32.

[59] Jong MC, Boers I, van Wietmarschen HA, Tromp E, Busari JO, Wennekes R, Snoeck I, Bekhof J, Vlieger AM. Hypnotherapy or transcendental meditation versus progressive muscle relaxation exercises in the treatment of children with primary headaches: a multi-centre, pragmatic, randomised clinical study. Eur J Pediatr 2019;178:147–154.

[60] Kipping K, Maier C, Bussemas HH, Schwarzer A. Medication compliance in patients with chronic pain. Pain Physician 2014;17:81–94.

[61] Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016;15:155–163.

[62] Koppenaal T, Linmans J, Knottnerus JA, Spigt M. Pragmatic vs. explanatory: An adaptation of the PRECIS tool helps to judge the applicability of systematic reviews for daily practice. J Clin Epidemiol 2011;64:1095–1101.

[63] Kortekangas T, Haapasalo H, Flinkkilä T, Ohtonen P, Nortunen S, Laine H-J, Järvinen TL, Pakarinen H. Three week versus six week immobilisation for stable Weber B type ankle fractures: randomised, multicentre, non-inferiority clinical trial. BMJ 2019;364:k5432.

[64] Krebs EE, Gravely A, Nugent S, Jensen AC, DeRonne B, Goldsmith ES, Kroenke K, Bair MJ, Noorbaloochi S. Effect of Opioid vs Nonopioid Medications on Pain-Related Function in Patients With Chronic Back Pain or Hip or Knee Osteoarthritis Pain: The SPACE Randomized Clinical Trial. JAMA 2018;319:872–882.

[65] Krebs EE, Lorenz KA, Bair MJ, Damush TM, Wu J, Sutherland JM, Asch SM, Kroenke K. Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. J Gen Intern Med 2009;24:733–738.

[66] Kroenke K, Baye F, Lourens SG, Evans E, Weitlauf S, McCalley S, Porter B, Matthias MS, Bair MJ. Automated Self-management (ASM) vs. ASM-Enhanced Collaborative Care for Chronic Pain and Mood Symptoms: the CAMMPS Randomized Clinical Trial. J Gen Intern Med 2019;34:1806–1814.

[67] Laughlin-Tommaso S, Barnard EP, AbdElmagied AM, Vaughan LE, Weaver AL, Hesley GK, Woodrum DA, Jacoby VL, Kohi MP, Price TM, Nieves A, Miller MJ, Borah BJ, Moriarty JP, Gorny KR, Leppert PC, Severson AL, Lemens MA, Stewart EA. FIRSTT study: randomized controlled trial of uterine artery embolization vs focused ultrasound surgery. Am J Obstet Gynecol 2019;220:174.e1-174.e13.

[68] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147.

[69] Loudon K, Zwarenstein M, Sullivan FM, Donnan PT, Gágyor I, Hobbelen HJSM, Althabe F, Krishnan JA, Treweek S. The PRECIS-2 tool has good interrater reliability and modest discriminant validity. J Clin Epidemiol 2017;88:113–121.

[70] Luoma KA, Leavitt IM, Marrs JC, Nederveld AL, Regensteiner JG, Dunn AL, Glasgow RE, Huebschmann AG. How can clinical practices pragmatically increase physical activity for patients with type 2 diabetes? A systematic review. Transl Behav Med 2017;7:751–772.

[71] Mansell NS, Rhon DI, Meyer J, Slevin JM, Marchant BG. Arthroscopic Surgery or Physical Therapy for Patients With Femoroacetabular Impingement Syndrome: A

Randomized Controlled Trial With 2-Year Follow-up. Am J Sports Med 2018;46:1306–1314.

[72] Masoudi FA, Havranek EP, Wolfe P, Gross CP, Rathore SS, Steiner JF, Ordin DL, Krumholz HM. Most hospitalized older persons do not meet the enrollment criteria for clinical trials in heart failure. Am Heart J 2003;146:250–257.

[73] McCarthy DM, Waite KR, Curtis LM, Engel KG, Baker DW, Wolf MS. What Did the Doctor Say? Health Literacy and Recall of Medical Instructions. Med Care 2012;50:277–282.

[74] McKee MD, Nielsen A, Anderson B, Chuang E, Connolly M, Gao Q, Gil EN, Lechuga C, Kim M, Naqvi H, Kligler B. Individual vs. Group Delivery of Acupuncture Therapy for Chronic Musculoskeletal Pain in Urban Primary Care-a Randomized Trial. J Gen Intern Med 2020;35:1227–1237.

[75] Mellor R, Bennell K, Grimaldi A, Nicolson P, Kasza J, Hodges P, Wajswelner H, Vicenzino B. Education plus exercise versus corticosteroid injection use versus a wait and see approach on global outcome and pain from gluteal tendinopathy: prospective, single blinded, randomised clinical trial. BMJ 2018;361:k1662.

[76] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1.

[77] Montedori A, Bonacini MI, Casazza G, Luchetta ML, Duca P, Cozzolino F, Abraha I. Modified versus standard intention-to-treat reporting: Are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. Trials 2011;12:58.

[78] Moore RA, Derry S, McQuay HJ, Straube S, Aldington D, Wiffen P, Bell RF, Kalso E, Rowbotham MC. Clinical effectiveness: An approach to clinical trial design more relevant to clinical practice, acknowledging the importance of individual differences. PAIN® 2010;149:173–176.

[79] Movsas B, Moughan J, Owen J, Coia LR, Zelefsky MJ, Hanks G, Wilson JF. Who Enrolls Onto Clinical Oncology Trials? A Radiation Patterns of Care Study Analysis. Int J Radiat Oncol 2007;68:1145–1150.

[80] Nicholls SG, Zwarenstein M, Hey SP, Giraudeau B, Campbell MK, Taljaard M. The importance of decision intent within descriptions of pragmatic trials. J Clin Epidemiol 2020;125:30–37.

[81] Nicolian S, Butel T, Gambotti L, Durand M, Filipovic-Pierucci A, Mallet A, Kone M, Durand-Zaleski I, Dommergues M. Cost-effectiveness of acupuncture versus standard care for pelvic and low back pain in pregnancy: A randomized controlled trial. PloS One 2019;14:e0214195.

[82] Noll E, Shodhan S, Romeiser JL, Madariaga MC, Page C, Santangelo D, Guo X, Pryor AD, Gan TJ, Bennett-Guerrero E. Efficacy of acupressure on quality of recovery after surgery: Randomised controlled trial. Eur J Anaesthesiol 2019;36:557–565.

[83] Northwood AK, Vukovich MM, Beckman A, Walter JP, Josiah N, Hudak L, O'Donnell Burrows K, Letts JP, Danner CC. Intensive psychotherapy and case management for Karen refugees with major depression in primary care: a pragmatic randomized control trial. BMC Fam Pract 2020;21:17.

[84] Nøst TH, Steinsbekk A, Bratås O, Grønning K. Twelve-month effect of chronic pain self-management intervention delivered in an easily accessible primary healthcare service - a randomised controlled trial. BMC Health Serv Res 2018;18:1012.

[85] O'Brien KM, Wiggers J, Williams A, Campbell E, Hodder RK, Wolfenden L, Yoong SL, Robson EK, Haskins R, Kamper SJ, Rissel C, Williams CM. Telephone-based weight loss support for patients with knee osteoarthritis: a pragmatic randomised controlled trial. Osteoarthritis Cartilage 2018;26:485–494.

[86] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann T, Mulrow CD, Shamseer L, Moher D. Mapping of reporting guidance for systematic reviews and meta-analyses generated a comprehensive item bank for future reporting guidelines. J Clin Epidemiol 2020;118:60–68.

[87] Palmer AJR, Ayyar Gupta V, Fernquest S, Rombach I, Dutton SJ, Mansour R, Wood S, Khanduja V, Pollard TCB, McCaskie AW, Barker KL, Andrade TJMD, Carr AJ, Beard DJ, Glyn-Jones S, FAIT Study Group. Arthroscopic hip surgery compared with physiotherapy and activity modification for the treatment of symptomatic femoroacetabular impingement: multicentre randomised controlled trial. BMJ 2019;364:l185.

[88] Park S-Y, Hwang E-H, Cho J-H, Kim K-W, Ha I-H, Kim M-R, Nam K, Lee MH, Lee J-H, Kim N, Shin B-C. Comparative Effectiveness of Chuna Manipulative Therapy for Non-Acute Lower Back Pain: A Multi-Center, Pragmatic, Randomized Controlled Trial. J Clin Med 2020;9.

[89] Persaud N, Mamdani MM. External validity: the neglected dimension in evidence ranking. J Eval Clin Pract 2006;12:450–453.

[90] Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Pearson/Prentice Hall Upper Saddle River, NJ, 2009 p.

[91] Qi Z, Pang Y, Lin L, Zhang B, Shao J, Liu X, Zhang X. Acupuncture Combined with Hydrotherapy in Diabetes Patients with Mild Lower-Extremity Arterial Disease: A Prospective, Randomized, Nonblinded Clinical Study. Med Sci Monit Int Med J Exp Clin Res 2018;24:2887–2900.

[92] Richmond H, Hall AM, Copsey B, Hansen Z, Williamson E, Hoxey-Thomas N, Cooper Z, Lamb SE. The Effectiveness of Cognitive Behavioural Treatment for Non-Specific Low Back Pain: A Systematic Review and Meta-Analysis. PLOS ONE 2015;10:e0134192.

[93] Riddle DL, Keefe FJ, Ang DC, Slover J, Jensen MP, Bair MJ, Kroenke K, Perera RA, Reed SD, McKee D, Dumenci L. Pain Coping Skills Training for Patients Who Catastrophize About Pain Prior to Knee Arthroplasty: A Multisite Randomized Clinical Trial. J Bone Joint Surg Am 2019;101:218–227.

[94] Rigoard P, Basu S, Desai M, Taylor R, Annemans L, Tan Y, Johnson MJ, Van den Abeele C, North R. Multicolumn spinal cord stimulation for predominant back pain in failed back surgery syndrome patients: a multicenter randomized controlled trial. Pain 2019;160:1410–1420.

[95] Roddy E-//-C. Open-label randomised pragmatic trial (CONTACT) comparing naproxen and low-dose colchicine for the treatment of gout flares in primary care. Ann Rheum Dis 2019.

[96] Roland M, Torgerson DJ. Understanding controlled trials: What are pragmatic trials? BMJ 1998;316:285.

[97] Rothrock JF, Adams AM, Lipton RB, Silberstein SD, Jo E, Zhao X, Blumenfeld AM, FORWARD Study investigative group. FORWARD Study: Evaluating the Comparative Effectiveness of OnabotulinumtoxinA and Topiramate for Headache Prevention in Adults With Chronic Migraine. Headache 2019;59:1700–1713.

[98] Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" The Lancet 2005;365:82–93.

[99] Rothwell PM, Watson CPN, Castro-Lopez J. External validity of randomized controlled trials: general principles and lessons from trials in neuropathic pain. Current topics in pain: 12th World Congress on Pain. IASP Press, Seattle, WA, 2009. pp. 199–220.

[100] Rowbotham MC, Gilron I, Glazer C, Rice ASC, Smith BH, Stewart WF, Wasan AD. Can pragmatic trials help us better understand chronic pain and improve treatment? PAIN 2013;154:643.

[101] Schneider MJ, Ammendolia C, Murphy DR, Glick RM, Hile E, Tudorascu DL, Morton SC, Smith C, Patterson CG, Piva SR. Comparative Clinical Effectiveness of Nonsurgical Treatment Methods in Patients With Lumbar Spinal Stenosis: A Randomized Clinical Trial. JAMA Netw Open 2019;2:e186828.

[102] Schulz C, Evans R, Maiers M, Schulz K, Leininger B, Bronfort G. Spinal manipulative therapy and exercise for older adults with chronic low back pain: a randomized clinical trial. Chiropr Man Ther 2019;27:21.

[103] Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis 1967;20:637–648.

[104] Semprini A, Singer J, Braithwaite I, Shortt N, Thayabaran D, McConnell M, Weatherall M, Beasley R. Kanuka honey versus aciclovir for the topical treatment of herpes simplex labialis: a randomised controlled trial. BMJ Open 2019;9:e026201.

[105] Sepehrvand N, Alemayehu W, Das D, Gupta AK, Gouda P, Ghimire A, Du AX, Hatami S, Babadagli HE, Verma S, Kashour Z, Ezekowitz JA. Trends in the Explanatory or Pragmatic Nature of Cardiovascular Clinical Trials over 2 Decades. JAMA Cardiol 2019;4:1122–1128.

[106] Smith DH, Kuntz JL, DeBar LL, Mesa J, Yang X, Schneider J, Petrik A, Reese K, Thorsness LA, Boardman D, Johnson ES. A randomized, pragmatic, pharmacist-led intervention reduced opioids following orthopedic surgery. Am J Manag Care 2018;24:515–521.

[107] Spall HGCV, Toren A, Kiss A, Fowler RA. Eligibility Criteria of Randomized Controlled Trials Published in High-Impact General Medical Journals: A Systematic Sampling Review. JAMA 2007;297:1233–1240.

[108] Sparv D, Hofmann R, Gunnarsson A, James S, Hedberg C, Lauermann J, Torild P, Omerovic E, Bergström K, Haugen E, Bergström C, Linder R, Borg P, Haaga U, Olsson A, Böving E, Östlund O, Rylance R, Witt N, Erlinge D, DETO2X-SWEDEHEART Investigators. The Analgesic Effect of Oxygen in Suspected Acute Myocardial Infarction: A Substudy of the DETO2X-AMI Trial. JACC Cardiovasc Interv 2018;11:1590–1597.

[109] Spiegel B, Fuller G, Lopez M, Dupuy T, Noah B, Howard A, Albert M, Tashjian V, Lam R, Ahn J, Dailey F, Rosen BT, Vrahas M, Little M, Garlich J, Dzubur E, IsHak

W, Danovitch I. Virtual reality for management of pain in hospitalized patients: A randomized comparative effectiveness trial. PloS One 2019;14:e0219115.

[110] Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng H-Y, Corbett MS, Eldridge SM, Emberson JR, Hernán MA, Hopewell S, Hróbjartsson A, Junqueira DR, Jüni P, Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling K, White IR, Whiting PF, Higgins JPT. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898.

[111] Stouten V, Westhovens R, Pazmino S, De Cock D, Van der Elst K, Joly J, Verschueren P, CareRA study group. Effectiveness of different combinations of DMARDs and glucocorticoid bridging in early rheumatoid arthritis: two-year results of CareRA. Rheumatol Oxf Engl 2019;58:2284–2294.

[112] Taylor AM, Phillips K, Patel KV, Turk DC, Dworkin RH, Beaton D, Clauw DJ, Gignac MAM, Markman JD, Williams DA, Bujanover S, Burke LB, Carr DB, Choy EH, Conaghan PG, Cowan P, Farrar JT, Freeman R, Gewandter J, Gilron I, Goli V, Gover TD, Haddox JD, Kerns RD, Kopecky EA, Lee DA, Malamut R, Mease P, Rappaport BA, Simon LS, Singh JA, Smith SM, Strand V, Tugwell P, Vanhove GF, Veasley C, Walco GA, Wasan AD, Witter J. Assessment of physical function and participation in chronic pain clinical trials: IMMPACT/OMERACT recommendations. PAIN 2016;157:1836.

[113] Thomson OP, Petty NJ, Moore AP. A qualitative grounded theory study of the conceptions of clinical practice in osteopathy – A continuum from technical rationality to professional artistry. Man Ther 2014;19:37–43.

[114] Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, Tunis S, Bergel E, Harvey I, Magid DJ, Chalkidou K. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol 2009;62:464–475.

[115] Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials 2009;10:37.

[116] Ulrich CM, Wallen GR, Feister A, Grady C. Respondent Burden in Clinical Research: When Are We Asking Too Much of Subjects? IRB Ethics Hum Res 2005;27:17–20.

[117] Verra ML, Angst F, Brioschi R, Lehmann S, Benz T, Aeschlimann A, De Bie RA, Staal JB. Effectiveness of subgroup-specific pain rehabilitation: a randomized controlled trial in patients with chronic back pain. Eur J Phys Rehabil Med 2018;54:358–370.

[118] Wang C, Schmid CH, Fielding RA, Harvey WF, Reid KF, Price LL, Driban JB, Kalish R, Rones R, McAlindon T. Effect of tai chi versus aerobic exercise for fibromyalgia: comparative effectiveness randomized controlled trial. BMJ 2018;360:k851.

[119] Wang Y, Lombard C, Hussain SM, Harrison C, Kozica S, Brady SRE, Teede H, Cicuttini FM. Effect of a low-intensity, self-management lifestyle intervention on knee pain in community-based young to middle-aged rural women: a cluster randomised controlled trial. Arthritis Res Ther 2018;20:74.

[120] Weinfurt KP. Clarifying the Meaning of Clinically Meaningful Benefit in Clinical Research: Noticeable Change vs Valuable Change. JAMA 2019;322:2381–2382.

[121] Whitty CJM, MacEwen C, Goddard A, Alderson D, Marshall M, Calderwood C, Atherton F, McBride M, Atherton J, Stokes-Lampard H, Reid W, Powis S, Marx C. Rising to the challenge of multimorbidity. BMJ 2020;368:l6964.

[122] Wibault J, Öberg B, Dedering Å, Löfgren H, Zsigmond P, Peolsson A. Structured postoperative physiotherapy in patients with cervical radiculopathy: 6-month outcomes of a randomized clinical trial. J Neurosurg Spine 2018;28:1–9.

[123] Williams A-//-W. Effectiveness of a healthy lifestyle intervention for chronic low back pain: a randomised controlled trial. Pain 03043959 2018;159:1137–1146.

[124] Williams HC, Burden-Teh E, Nunn AJ. What Is a Pragmatic Clinical Trial? J Invest Dermatol 2015;135:1–3.

[125] Witt CM, Manheimer E, Hammerschlag R, Lüdtke R, Lao L, Tunis SR, Berman BM. How Well Do Randomized Trials Inform Decision Making: Systematic Review Using Comparative Effectiveness Research Measures on Acupuncture for Back Pain. PLOS ONE 2012;7:e32399.

[126] World Health Organization. WHO glossary: Clinical trials. World Health Organ n.d. Available: https://www.who.int/topics/clinical_trials/en/. Accessed 5 Mar 2020.

[127] Zhuang Q, Tao L, Lin J, Jin J, Qian W, Bian Y, Li Y, Dong Y, Peng H, Li Y, Fan Y, Wang W, Feng B, Gao N, Sun T, Lin J, Zhang M, Yan S, Shen B, Pei F, Weng X. Postoperative intravenous parecoxib sodium followed by oral celecoxib post total knee arthroplasty in osteoarthritis patients (PIPFORCE): a multicentre, double-blind, randomised, placebo-controlled trial. BMJ Open 2020;10:e030501.

[128] Zuidgeest MG, Goetz I, Grobbee DE. PRECIS-2 in perspective: what is next for pragmatic trials? J Clin Epidemiol 2017;84:22–24.

[129] Zwarenstein M, Thorpe K, Treweek S, Loudon K. PRECIS-2 for retrospective assessment of RCTs in systematic reviews. J Clin Epidemiol 2020;126:202–206.

[130] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D, CONSORT group, Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008;337:a2390.

[131] Zwarenstein M, Treweek S, Loudon K. PRECIS-2 helps researchers design more applicable RCTs while CONSORT Extension for Pragmatic Trials helps knowledge users decide whether to apply them. J Clin Epidemiol 2017;84:27–29.

Figure legends

Figure 1: PRISMA flow diagram *showing the identification, screening and selection process of records for the present systematic review. Reasons for exclusion at the full-text eligibility-screening phase are provided.*

Figure 2: Average PRECIS-2 scores per domain for all included trials. *Standard deviations and n not indicated (see supplementary table*, available at http://links.lww.com/PAIN/B374*). Less pragmatic design choices result in 'dents' in the*

wheel diagram whilst higher average ratings per domain cause the line to be closer towards the rim of the wheel.

Table 1: Treatment modalities and demographics.

Component therapeutic modalities	**n of trials**	**%**	**Trial references**
Pharmacological therapy	12	21.05	[5,20,24,37,45,66,97,99,106,110,113,129]
Cognitive-behavioural & other psychotherapy	9	15.79	[8,17,68,85–87,95,121,125]
Surgery	7	12.28	[6,7,51,69,73,89,96]
Acupuncture / acupressure	6	10.53	[10,21,76,83,84,93]
Manual Therapy	6	10.53	[31,36,50,52,90,104]
Physiotherapy	4	7.02	[9,103,119,124]
Multidisciplinary care (non-drug)	3	5.26	[3,27,35][3][2] [3]
General Practice (non-drug)	2	3.51	[19,30]
Rehabilitation	2	3.51	[65,77]
Body-mind therapies	2	3.51	[61,120]
Education	1	1.75	[108]
Automated symptom and treatment side effect monitoring	1	1.75	[2]
Virtual reality	1	1.75	[111]
Dentistry	1	1.75	[40]
Pain disorder / descriptor	**n of trials**	**%**	**Trial references**
Back or neck pain	19	33.33	[19,24,27,30,31,36,45,50,52,66,76,83,90,96,103,104,119,124,125]
Peripheral joint pain	10	17.54	[7,35,51,73,77,87,89,95,97,121]

Arthritis (RA and / or OA)	8	14.04	[6,8,35,37,66,76,87,113]
Pain (not further specified)	6	10.53	[9,17,68,85,86,111]
Post-medical intervention pain	5	8.77	[84,96,108,124,129]
Abdominal and other visceral pain	4	7.02	[10,21,69,110]
Neuropathic pain	3	5.26	[2,20,106]
Headaches	3	5.26	[5,61,99]
Leg pain	2	3.51	[93,103]
Post-injury pain	2	3.51	[27,65]
Tooth pain	1	1.75	[40]
Diffuse chronic pain (CFS, FM, CRPS)	1	1.75	[120]
Musculoskeletal pain (not further specified)	1	1.75	[3]
Pain duration	**n of trials**	**%**	**Trial references**
Acute	7	12.28	[27,30,45,65,84,106,110]
Subacute	2	3.51	[37,83]
Chronic	31	54.39	All others
Mixed	1	1.75	[50]
Not reported	16	28.07	[2,5,9,10,19,31,35,36,40,69,73, 85,89,108,111,121]
Type of setting	**n of trials**	**%**	**Trial references**
Primary	25	43.86	*Not provided, see below for detail*
Secondary	20	35.09	
Tertiary	17	29.82	
Community	5	8.77	
Setting specification	**n of trials**	**%**	**Trial references**
Public hospital	23	40.35	All others, unclear in: [83,96]

Private hospital	6	10.53	[73,90,93,96,99,119]
Patient home, phone, text messaging, mail, or online (entirely or predominantly)	6	10.53	[17,68,87,104,108,124]
Private practice	8	14.04	[9,31,36,52,76,77,99,113]
Military medical practice	3	5.26	[24,50,66]
Research institute	1	1.75	[104]
Rehabilitation centre	1	1.75	[8]
Emergency Dept.	1	1.75	[45]
University teaching clinic	1	1.75	[21]

All therapies studied across included trials as well as pain disorders, and the average duration of pain are presented. Note: Some pain descriptors have been applied twice, e.g., knee arthritis has been classified as 'Peripheral joint pain' and 'Arthritis'. Further, some samples included patients with musculoskeletal pain, which may have included 'Back or neck pain' and 'peripheral joint pain'. Depending on individual trial reporting, the average duration of the pain-related diagnosis or the time since the onset of pain was used. Acute pain was defined as pain lasting < 4 weeks, subacute as 4 weeks to 3 months, and chronic as 3 months. CFS, chronic fatigue syndrome; CRPS, chronic regional pain syndrome; FM, fibromyalgia; RA, rheumatoid arthritis; OA, osteoarthritis.

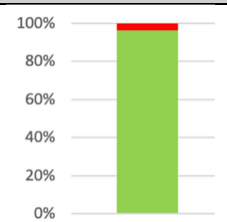
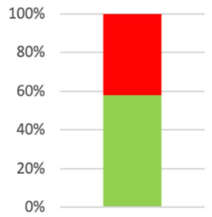
Table 2: Methods of trial design.

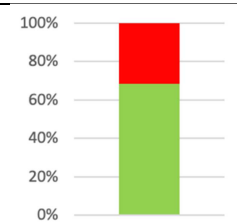
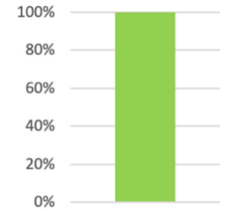
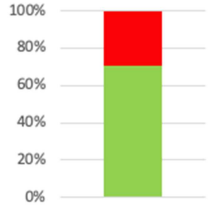
Comparator	**n of trials**	**% of sample**	**Trial references**
Another active specific therapy (comparative effectiveness) *	29	50.88	All others
Treatment / care as usual *	14	24.56	[3,8,10,17,30,35,50,83–85,87,95,108,124]
Placebo or sham intervention	5	8.77	[5,7,45,84,129]
no treatment group (explicitly assigned, i.e., patient know they won't get any treatment)	4	7.02	[7,9,93,110]
Treatment / care as usual plus something else (e.g., advice, education, etc.) *	2	3.51	[2,93]
Waitlist control	1	1.75	[125]
Advice only	1	1.75	[121]
Wait & see (not waitlist but monitoring)	1	1.75	[77]
No-treatment group (but unaware of trial)	1	1.75	[19]
Recruitment method			
Targeted recruitment (such as identification through records)	25	43.86	All others
Convenience sampling	16	28.07	[5,6,8,9,20,27,30,40,45,52,65,73,76,89,95,110]
Not reported	9	15.79	[24,37,61,93,111,113,119,124,129]
Mixed (convenience and targeted)	8	14.04	[35,50,51,69,83,86,96,97]

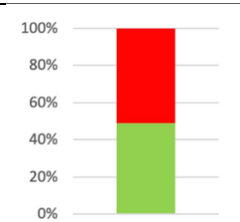
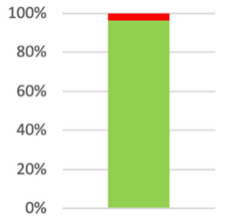
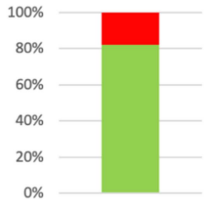
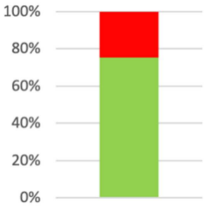
Method of randomization			
individually randomized	27	47.37	
of which simple randomization	15	26.32	All others
of which blocked randomization	12	21.05	[17,21,27,31,36,37,45,68,73,90,104,119]
stratified by site	15	26.32	[5,7,20,24,50,51,61,65,69,76,77,83,96,106,129]
other stratification	9	15.79	[66,84,86,89,93,95,103,113,120]
Cluster randomised	6	10.53	[2,3,8,30,35,121]

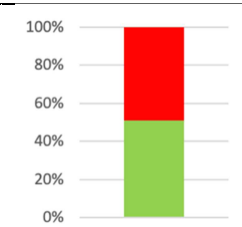
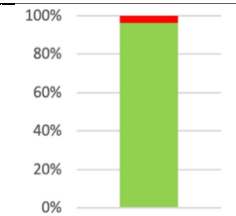
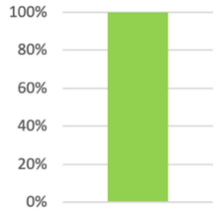
*Employed comparators and recruitment and randomization methods are presented. Notes: Multiple comparator groups were possible. The difference between a waitlist control group and a no-treatment control group is that patients expect treatment at a later point or know that they have been assigned to not receiving any treatment, respectively. Categories marked * are deemed part of Comparative Effectiveness Research (CER). Convenience sampling is the recruitment of patients who attend the trial-delivering service anyway, whilst targeted strategies seek to specifically contact populations of potentially eligible participants. The category of 'blocked randomization' includes various was of blocking, including a single fixed block size, regularly varying sizes, and randomly permuted block sizes. Blocking was occasionally stratified by site. Stratification was usually by trial centres (sites). 'Other stratification' includes stratification by gender, diagnosis or treating surgeon. Cluster randomization refers to trials where the unit of randomization were not patients but, for example, clinics or individual providers.*

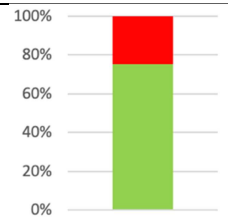
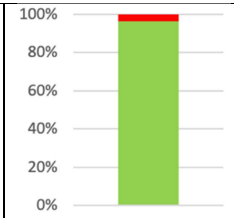
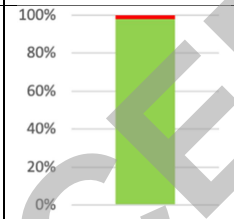
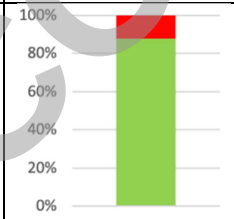
Table 3: Selected items of the 2010 update of the Consolidated Standards of Reporting Trials (CONSORT) statement (Schulz et al., 2010)[1] and all items of the extension for the reporting of pragmatic trials, as published in 2008 by Zwarenstein et al.[2].

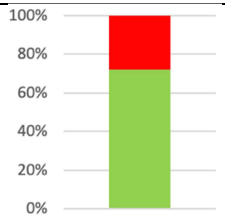
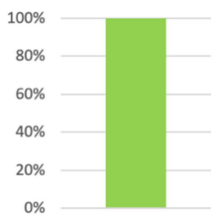
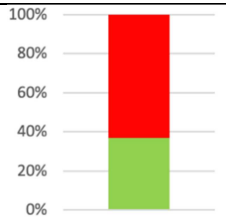
Item (number, section & CONSORT document (1 or 2))	Description of reporting item	Results: Number of studies which complied with respective reporting items (n, %)
2: Background[2] modified	Describe the health or health service problem that the intervention is intended to address	55 (96.49%) Not complied: [6,110] 
	and other interventions that may commonly be aimed at this problem modified	33 (57.89%) 
3: Participants[2]	Eligibility criteria should be explicitly framed to show the degree to which they include typical participants **and/or, where applicable,** typical providers (e.g., nurses), institutions (e.g., hospitals), communities (or	39 (68.42%)

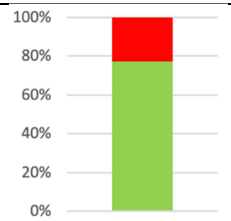
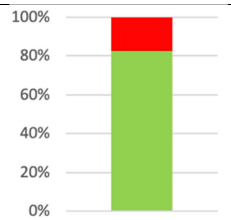
	localities e.g., towns) and settings of care (e.g., different healthcare financing systems)	
5: Interventions[1] modified	Precise details of the interventions intended for the intervention group ~~each group and how and when they were actually administered~~	57 (100%) 
	If the 'treatment-as-usual' or 'usual care' was used as comparator, provide additional information as to the nature of the intervention(s) available as part of this	(applicable in n = 17) 12 (70.59%) 
	If the 'treatment-as-usual' or 'usual care' was employed as comparator, collect and report data on care received by patients in this group	(applicable in n = 17) 10 (58.82%)

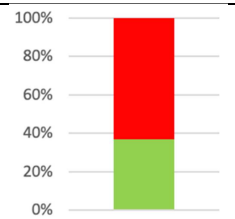
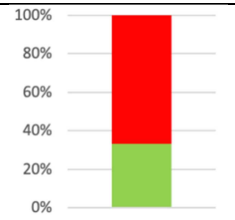
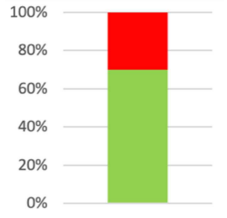
4 Interventions[2]	Describe extra resources added to (or resources removed from) usual settings in order to implement intervention	28 (49.12%) 
	Describe the health or health service problem that the intervention is intended to address	55 (96.49%) 
	Indicate if efforts were made to standardize the intervention or if the intervention and its delivery were allowed to vary between participants, practitioners, or study sites	Not applicable as intervention automated: 1; Reported: 46 (82.14% of 56); of those standardized: 35; Not standardized: 9. 
	Describe the comparator in similar detail to the intervention	43 (75.44%) 

6 Outcomes[2]	Explain why the chosen outcomes and, when relevant, the length of follow-up are considered important to those who will use the results of the trial	29 (50.88%) 
7a: Sample size[1]	Report how the sample size was determined.	55 (96.49%) 
7: Sample size[2]	If calculated using the smallest difference considered important by the target decision maker audience (the minimally important difference) then report where this difference was obtained	*Not extracted*
8b: Randomization[1]	Report the type of randomization & details of any restriction (such as blocking and block size).	57 (100%) 

10: Allocation concealment implementation	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions?		43 (75.44%) 
11a: Blinding / masking[1] modified	Whether participants,		Not applicable in one case as patients unaware of participating in a trial [105]. Reported in 54 (96.43%) of 56; Not reported in [65,82]
	those administering the interventions,		Not applicable in one case as intervention independent of providers. Reported in 55 (98.2%) of 56 relevant trials.
	and those assessing the outcomes were blinded to group assignment		Reported in 50 cases (87.72%).

11: Blinding / masking[2]	If blinding was not done, or was not possible, explain why.		31 (72%) of relevant studies reported reasons (n = 43).
13a: Participant flow[1] modified	Flow of participants through each stage (a diagram is strongly recommended)—specifically, for each group, report the numbers of participants randomly assigned		57 (100%) 
	~~receiving intended treatment~~		*Not extracted*
	completing the study protocol		57 (100%)
	analysed for the primary outcome;		57 (100%)
	describe deviations from planned study protocol, together with reasons		21 (36.84%) complied. Of those: 10 reported following protocol; Deviations with reasons reported in 11; Deviated without providing reasons: 4. 

13: Participant flow[2]	The number of participants or units approached to take part in the trial, the number which were eligible, and reasons for non-participation should be reported.	44 (77.19%) 
16: Numbers analyzed[1] modified	Whether the analysis was by 'intention-to-treat'; For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	48 (84.21%) reported the primary analysis as 'Intention-to-treat'; Not discernible in one instance [18]; all other did not explicitly report a primary Intention-to-treat analysis. Out of those 48 studies which called their primary analysis 'Intention-to-treat', ten trials (20.83%) excluded participants from the primary analysis who did not provide follow-up data or where data were missing.
19: Harms[1]	All important harms or unintended effects in each group.	Whether or not significant harms or unintended effects occurred was reported in 47 studies (82.46%).  Harms did occur in 22 of those studies, and in nine of those cases there was a significant difference between groups [23,62,70,80,83,93,94,96,126].

21: Generalizability[2] modified	Describe key aspects of the setting which determined the trial results.	21 (36.84%) 
	Discuss possible differences in other settings where clinical traditions, health service organisation, staffing, or resources may vary from those of the trial	19 (33.33%) 
25: Funding[1]	Sources of funding and other support (such as supply of drugs) **and** role of funders.	40 (70.18%) 

Where indicated by colour or strikethrough formatting, items were modified to match the review question or split into several items in order to be extractable as individual data points. Bar charts indicate the percentage of studies complying with respective reporting items (green) and not complying even though applicable (red). If items were not applicable to the entire sample of 57 studies, the applicable number is stated in the respective row.

Table 4: Correlation analysis between domain-specific PRECIS-2 ratings and total sample size at randomization.

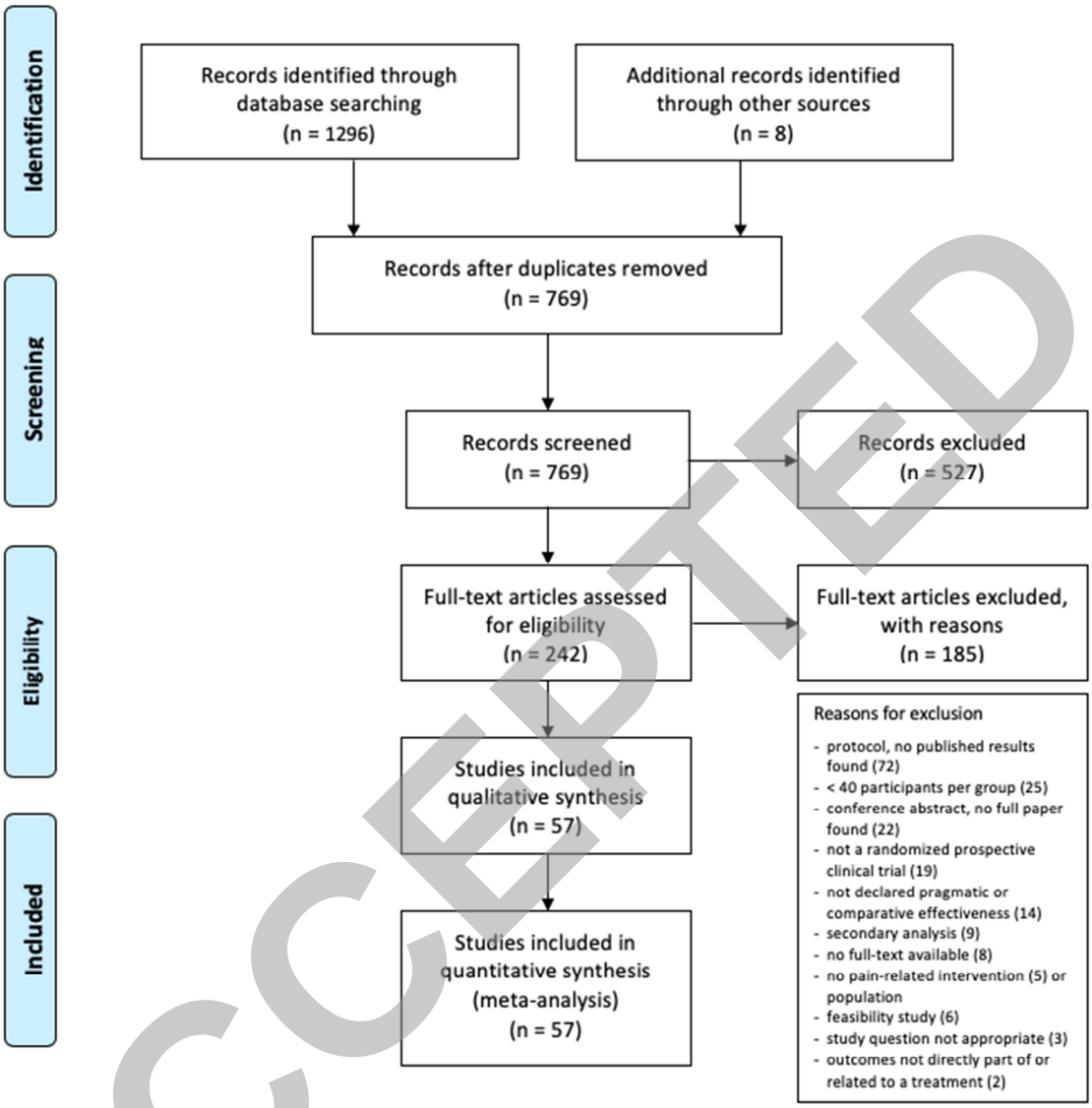
	PRECIS-2 domain	Eligibility	Recruitment	Setting	Organization	Flexibility (delivery)	Flexibility (adherence)	Follow-up	Outcome	Analysis	PRECIS-2 score	Different. attrition
Total sample size at BL	Correl. coeff.	.250	.029	.451**	.281*	.031	.193	.190	.161	.273*	.408**	-.360**
	p	.061	.846	.000	.046	.822	.155	.157	.232	.040	.002	.006
	n	57	47	56	51	56	56	57	57	57	57	57

*Also showing the correlation analysis between sample size and differential attrition (last column) as measured by the difference in drop-outs between irrespective of the 'direction' of attrition, i.e., in which group more patients were lost to follow-up (Spearman's Rank Order Correlation used as data not conforming with normality assumption; * $p < 0,05$, ** $p < 0,01$; 2-tailed). BL, baseline.*

Table 5: Correlation analyses amongst trial methods with ratings of trial pragmatism (PRECIS-2 scores), randomization methods, and analysis method.

	DV	Results	Sensitivity analysis
PRECIS-2 average	Sample size		
	Number of trial centres †	.190, p = .191, (df = 47)	
	Funding source †	-.048, p = 727, (df = 54)	
	Index therapy †	-.005, p = .97, (df = 54)	.089, p = .512, n = 57
	Index pain disorder †	-.285*, p = .033, (df = 54)	-.213, p = .112, n = 57
	Analysis method †	.198, p = .16, (df = 50)	.284*, p = .039, n = 53
Randomization method	PRECIS-2 average †	-.197, p = .145, (df = 54)	-.137, p = .31, n = 57
	Baseline heterogeneity	*Analysis not conducted as no trials with sign. between-group age differences at baseline*	
	Sample size	.21, p = .122, n = 57	
	Analysis method	.0, p = .99, n = 53	
	Funding source	-.072, p = .594, n = 57	
Blinding of participants	PRECIS-2 average †	.214, p = .124 (df = 51)	.135, p = .331, n = 54
	Sample size	-.081, p = .562, (n = 54)	
	Analysis method	.2, p = .16, (n = 51)	
	Funding source	.111, p = .424, (n = 54)	

*Statistical tests were part correlation analyses where covariates were controlled for and Spearman's rho where this was not indicated. † Sample size was used as covariate of no interest. Sensitivity analyses assess the same correlation without controlling for pre-identified confounding variables. * significant at $p < 0.05$ level (2-tailed). DV, dependent variable.*


Identification
Screening
Eligibility
Included
Records identified through database searching (n = 1296)
Additional records identified through other sources (n = 8)
Records after duplicates removed (n = 769)
Records screened (n = 769)
Records excluded (n = 527)
Full-text articles assessed for eligibility (n = 242)
Full-text articles excluded, with reasons (n = 185)
Reasons for exclusion
- protocol, no published results found (72)
- < 40 participants per group (25)
- conference abstract, no full paper found (22)
- not a randomized prospective clinical trial (19)
- not declared pragmatic or comparative effectiveness (14)
- secondary analysis (9)
- no full-text available (8)
- no pain-related intervention (5) or population
- feasibility study (6)
- study question not appropriate (3)
- outcomes not directly part of or related to a treatment (2)
Studies included in qualitative synthesis (n = 57)
Studies included in quantitative synthesis (meta-analysis) (n = 57)


Eligibility
5
4
3
2
1
Recruitment
Setting
Organisation
Flexibility (delivery)
Flexibility (adherence)
Follow-up
Primary outcome
Primary analysis