

## REVIEW

# Analgesic efficacy of opioids in chronic pain: recent meta-analyses

H Reinecke<sup>1</sup>, C Weber<sup>1</sup>, K Lange<sup>1</sup>, M Simon<sup>1</sup>, C Stein<sup>2</sup> and H Sorgatz<sup>1</sup><sup>1</sup>Department of Psychology, Technische Universität Darmstadt, Darmstadt, Germany, and<sup>2</sup>Department of Anesthesiology and Critical Care Medicine, Freie Universität Berlin, Charité Campus Benjamin Franklin, Berlin, Germany**Correspondence**

Christoph Stein, Department of Anesthesiology and Critical Care Medicine, Freie Universität Berlin, Charité Campus Benjamin Franklin, Hindenburgdamm 30, D-12203 Berlin, Germany. E-mail: christoph.stein@charite.de

**Received**

15 October 2013

**Revised**

20 January 2014

**Accepted**

31 January 2014

Opioids are regularly administered in acute and cancer pain. In chronic non-cancer pain (CNCP), however, their use is controversial. Previous meta-analyses and randomized controlled trials (RCTs) lack methodological homogeneity and comparable data. Here we analysed the maximum analgesic efficacies of opioids and non-opioids compared with placebo, and of physiotherapy and psychotherapy compared with active or waiting-list controls. We screened 3647 citations and included RCTs if treatment duration was at least 3 weeks, data were sufficient for meta-analysis, and criteria for high quality were met. Only 46 studies (10 742 patients) met the criteria. Weighted and standardized mean differences (WMD, SMD) between pain intensities were pooled to conduct separate meta-analyses for each treatment category. At the end of treatment the WMD for pain reduction (100-point scale) was 12.0 for 'strong' opioids, 10.6 for 'weak' opioids, 8.4 for non-opioids (each vs. placebo), 5.5 for psychotherapy and 4.5 for physiotherapy (each vs. active controls). Dropout rates were high in pharmacological studies. The 95% confidence intervals using the outcomes of control groups did not indicate statistical differences between efficacies of the five interventions. Because not enough eligible head-to-head trials were available, our analysis is limited to adjusted indirect comparisons. The heterogeneity of pre-post pain differences in control groups did not allow the definition of a common comparator. In conclusion, although there were statistically significant differences between maximum treatment efficacies, no intervention *per se* produced clinically important improvements in average pain intensity. Thus, opioids alone are inappropriate and multimodal treatment programmes may be required for CNCP.

**LINKED ARTICLES**

This article is part of a themed section on Opioids: New Pathways to Functional Selectivity. To view the other articles in this section visit <http://dx.doi.org/10.1111/bph.2015.172.issue-2>

**Abbreviations**

AC, active control; CI, confidence interval; CNCP, chronic non-cancer pain; n.s., not significant; NSAID, nonsteroidal anti-inflammatory drug; RCT, randomized controlled trial; SIGN, Scottish Intercollegiate Guidelines Network; SMD, standardized mean difference; WHO, World Health Organization; WLC, waiting-list control; WMD, weighted mean difference

**Introduction**

Opioids are regularly administered for the treatment of acute and cancer pain. In chronic non-cancer pain (CNCP), however, their use is controversial. CNCP (e.g. arthritic, neuropathic, low back pain) is a result of complex bio-

psychosocial interactions with high socio-economic impact. Up to one-quarter of the general population and a significant portion of the workforce is affected (Breivik *et al.*, 2006; Solomon, 2010; Toblin *et al.*, 2011; Von Korff *et al.*, 2011; Pizzo and Clark, 2012), and huge market sales (up to  $50 \times 10^9$  US \$) of analgesic drugs are quoted (Melnikova, 2010).

Treatment aimed to keep pain intensity at a tolerable level and to improve physical and social functioning, behavioural patterns, quality of life and mood (Stein, 1997; 2013).

Long-term pharmacological treatment of CNCP is highly disputed. Non-opioids (e.g. nonsteroidal anti-inflammatory drugs, NSAIDs) can induce gastrointestinal ulcers, bleeding or cardiovascular complications because of ubiquitous COX inhibition (Woodcock, 2009; Trelle *et al.*, 2011), and opioids can produce cognitive impairment, respiratory depression, tolerance and addiction by activating opioid receptors in the brain (Stein *et al.*, 2003; Furlan *et al.*, 2006; Benyamin *et al.*, 2008; Von Korff *et al.*, 2011). Drastic increases in opioid prescriptions associated with diversion, abuse, overdose, and death have been reported (Martell *et al.*, 2007; Okie, 2010; Bohnert *et al.*, 2011; Von Korff *et al.*, 2011; Paulozzi *et al.*, 2012).

To rationalize these debates, it is imperative to scrutinize the efficacy of current drug treatments and to compare them to non-pharmacological approaches. Recent meta-analyses of randomized controlled trials (RCTs) on opioids versus placebo described maximum mean pain reductions of 14 units on a 100-point scale and standardized mean differences (SMDs) between 0.19 and 0.70 (Furlan *et al.*, 2006; Martell *et al.*, 2007; Papaleontiou *et al.*, 2010; Chaparro *et al.*, 2013; McNicol *et al.*, 2013). Larger effects were only reported in uncontrolled observational studies (Noble *et al.*, 2010; Gatti *et al.*, 2011). However, most RCTs merely used effect sizes at the end of treatment (post-post comparisons). Only one meta-analysis calculated differences between baseline and end of treatment (pre-post) from least square estimates for each study group (Papaleontiou *et al.*, 2010), but this may lead to over- or underestimation of effect sizes (Prakash *et al.*, 2008). In addition, none of the recent meta-analyses, reviews or guidelines numerically quote the efficacy of opioids relative to other therapies for CNCP (Chou *et al.*, 2009b; Management of Opioid Therapy for Chronic Pain Working Group, 2010; National Opioid Use Guideline Group, 2010), but some demand minimum average pain reductions (e.g. 30%) with opioid treatment (Management of Opioid Therapy for Chronic Pain Working Group, 2010; Turk *et al.*, 2011). So far, it is not possible to compare average values of pain decrease between opioid, non-opioid and non-pharmacological treatments because of the methodological heterogeneity of available studies. In particular, maximum achievable pain reductions calculated exclusively on the basis of high-quality RCTs are not known and conclusions regarding the clinical relevance of such values cannot be drawn.

To develop realistic treatment goals, maximum achievable efficacy values must be available. As for clinical decision-making, effect sizes must be considered in the context of alternative treatments, our objective was to analyse the best evidence on analgesic effectiveness of the three drug classes: non-opioids [World Health Organization (WHO) I], 'weak' (WHO II) and 'strong' opioids (WHO III), in relation to physiotherapy and psychotherapy. To determine the optimally achievable efficacy for each of these approaches, we conducted five independent, methodologically unified meta-analyses. By including only high-quality published RCTs, we tried to avoid fallacies from merging data of high quality with those of low-quality or unpublished trials (Chan and Altman, 2005; Kirkham *et al.*, 2010). The present analyses are an

extension of work initiated by a panel gathered by the German Society for the Study of Pain to develop clinical guidelines for the long-term use of opioids (<http://www.uni-duesseldorf.de/awmf/11/041-003.htm>). Parts of the data were discussed within this panel.

## Methods

### *Study characteristics, databases and search strategy*

We included all high-quality RCTs on CNCP except on headache because this was frequently reported as adverse side effect of opioid therapy (Papaleontiou *et al.*, 2010). Using the PRISMA guidelines (Moher *et al.*, 2009), we conducted five meta-analyses of studies on WHO I, II and III analgesics compared with placebo, and on physiotherapy and psychotherapy compared with active (AC) or waiting-list controls (WLC). The primary outcome had to be pain intensity measured with single scales linearly transformable to a 0- to 100-point scale such as the numerical rating scale or the visual analogue scale (Herr *et al.*, 2004).

We searched for published studies in PubMed, ClinicalTrials.gov, Cochrane Central Register of Controlled Trials (for analgesics and physiotherapy), the Oxford Pain Internet Site, PSYINDEX and PsycINFO (for psychotherapy), EMBASE, personal literature inventories, publisher databases and pharmaceutical companies without language restriction as detailed in the Supporting Information. As previous meta-analyses included publications since 1960, but found no eligible placebo-controlled studies on opioid analgesics before 1990 (Avouac *et al.*, 2007; Chou and Carson, 2008), we searched from 1990 to 2009. Only RCTs of at least 3 weeks duration were included because shorter trials used highly variable dosages during the titration phase.

### *Study selection and risk of bias in individual studies*

Abstracts of papers were screened for eligibility. If not excluded based on the abstract, full copies of the papers were obtained and screened again. Studies were included if numbers of patients, means and SDs for relevant treatment and control groups were reported. Two authors independently decided whether an individual RCT met the inclusion criteria. Two different reviewers then performed a qualitative evaluation of internal bias according to the Scottish Intercollegiate Guidelines Network (SIGN, 2011). Because double-blinding is difficult in non-pharmacological trials, we used an adapted version of the SIGN checklist for physiotherapy studies and a special grading system for psychotherapy studies (Yates *et al.*, 2005; Wissenschaftlicher Beirat Psychotherapie, 2007). If inter-rater reliabilities were lower than  $r = 0.80$ , the four reviewers discussed differing points and formed a consensus. The quality (level of evidence) was rated 1 for RCTs and distinguished by '++' (good, very low risk of bias), '+' (fair, low risk of bias) or '-' (poor, high risk of bias). Only studies rated '+' or '++' were included in the main meta-analyses. Meta-analyses of 'poor' rated studies were conducted if effect measures could be computed or imputed from published data.

### Data collection

Four authors (C. W., H. R., K. L., M. S.) independently extracted data using a standardized protocol. Each set was counterchecked by another author. Disagreements were settled by consensus. Extracted data included pain syndrome, characteristics of treatment and control groups, treatment duration, pain intensity scores (means, SD) at baseline (pre) and at the last point of measurement (post), number of patients (pre, post) and dropouts, functional outcomes (quality of life, quality of sleep, physical functioning), adverse events, use of rescue- and supplemental medication, and concomitant treatments. Pain scores deviating from a 100 point scale were transformed linearly to a 0-100 point scale (Herr *et al.*, 2004). In case of incomplete pain data (mean pre/post, differences) or number of patients (pre/post, dropouts), the missing values were computed by addition or subtraction of data presented in the same study.

### Summary measures

Using Cochrane Review Manager 5 ([Computer program], The Nordic Cochrane Centre, The Cochrane Collaboration, 2010, Copenhagen, Denmark), weighted mean differences (WMDs) and SMDs between post (end of treatment) pain scores of treatment and control groups, and between pre (baseline) and post scores of each group were calculated as the principal outcome measures. Statistical consistency was verified by performing a Q-test ( $\chi^2$ ) and reported as  $I^2$ -statistic (higher values mean higher heterogeneity). All analyses were calculated with the conservative random effects model (Iverson *et al.*, 2007). RCTs with multiple treatment groups were split into single intervention-control comparisons. Thus, a study with two different treatment groups was factored twice into the calculations. Analyses were conducted separately for studies on WHO I-, II- and III-drugs, physiotherapy and psychotherapy.

### Risk of bias across studies

To assess the impact of bias, an analysis was conducted combining high-quality (++, +) pharmacological trials with low-quality (–) studies (which were excluded from the main analyses). Publication bias was examined by estimating the small study effect for each treatment category using Egger's regression intercept (Comprehensive Meta-Analysis software; Biostat, Inc., Englewood, NJ, USA; Nuesch *et al.*, 2010).

### Additional analyses

For measures of functioning, RCTs reporting secondary outcomes in at least one of the following categories were identified: physical functioning, quality of sleep and health-related quality of life (Supporting Information Table S1). Dropouts were noted, including underlying reasons if mentioned, and risk differences were calculated for the prevalence of adverse effects in treatment and control groups. Subgroup analyses for pain syndromes and treatment durations were also performed.

## Results

### Study selection

A flow diagram is presented in Figure 1. A total of 3647 records were screened. Sixty-six pharmacological studies, 20

physiotherapy studies and 15 psychotherapy studies met the inclusion criteria and were evaluated for risk of bias. Methodological quality was adequate in 35 of 66 pharmacological studies, 9 of 20 physiotherapy studies and 13 of 15 psychotherapy studies. Twenty-four of the eligible pharmacological studies and all adequate physiotherapy and psychotherapy studies ( $n = 22$ ) reported data necessary for a meta-analysis. Thus, 46 high-quality studies (10 742 patients) constituted the final sample. All included trials are listed in the Supporting Information. Twenty-two of the included 24 pharmacological studies reported pain scores of an intention-to-treat or a predefined evaluable population, two studies presented only data of patients completing the study. A detailed quality assessment for each pharmacological study is presented in Supporting Information Table S2. Unpublished trials provided by pharmaceutical companies did not match our inclusion criteria because of the lack of control groups or blinding.

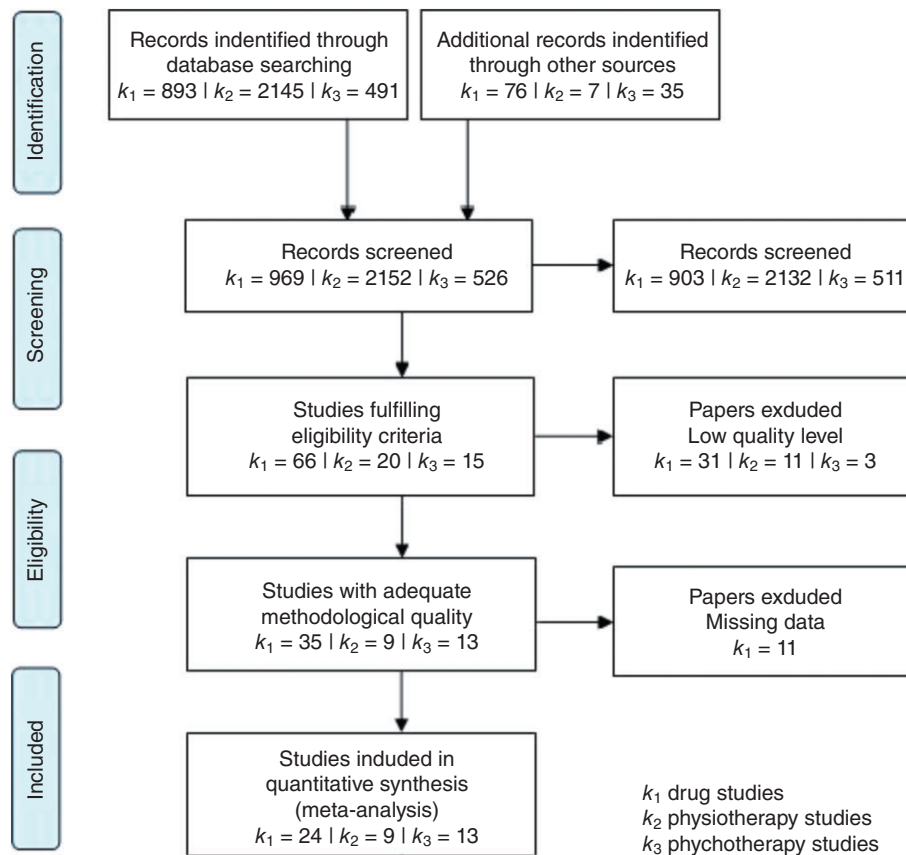
### Computation of missing data

Eleven of 35 pharmacological studies with otherwise adequate methodological quality outcomes were reported incompletely so that meta-analytic calculations were not possible without computations. In 7 of the remaining 24 studies, post SDs were not reported. As these omissions should not be classified as data 'missing at random', computation procedures as recommended by Cochrane might be misleading (Allison, 2003; Idris and Robertson, 2009). To include these seven studies (with over 5000 patients) into the analysis, the mean of the ratios ( $=1.33$ ) between post SD and baseline SD of included studies with complete data were used to impute missing post SDs from their respective baseline SDs. An alternative computation procedure on the basis of linear regression ( $a = -0.49$ ;  $b = 31$ ;  $R^2 = 0.16$ ) between published baseline SD and post SD resulted in comparable central parameters ( $SD_{\text{post mean}} = SD_{\text{post median}} = 23$ ), but reduced the range from 11.4 to 3.8, and from 15.2 to 5.3. The replacement of missing SDs by  $SD_{\text{pooled}}$  showed only small differences in SMDs ( $<0.03$ ) and WMDs ( $<0.33$ ) (Furukawa *et al.*, 2006). Because this replacement method has not been empirically examined for differences between pre and post scores, we did not use it.

### Study characteristics

Opioid studies comprised treatment of arthritis, low back pain, neuropathic pain and fibromyalgia with codeine or tramadol (WHO II), and with fentanyl, morphine, oxycodone or oxymorphone (WHO III) (see also Alexander *et al.*, 2013a). Studies investigating non-opioids (WHO I) included treatment of rheumatoid or osteoarthritis with acetaminophen (paracetamol), diclofenac, naproxen and various 'coxibs' (see also Alexander *et al.*, 2013b). Active placebos (benztropine, lorazepam) were used in only two WHO III crossover trials. As these studies did not provide detailed pain scores or numbers of patients per condition for the first phase, they were analysed separately. Mean baseline pain scores did not differ between WHO I, WHO II and WHO III studies (ANOVA,  $F = 1.94$ , d.f. = 2/20,  $P = 0.17$ ).

Physiotherapy studies comprised over 15 approaches (e.g. ultrasound, thermotherapy, nerve stimulation, etc.) and included patients with musculoskeletal (predominantly back-, neck- and/or shoulder pain), arthritic or neuropathic

**Figure 1**

Selection of studies.

pain. The AC groups in physiotherapy studies were treated with light aerobics, sham ultrasound or sham neurostimulation. Psychotherapy studies investigated cognitive and/or behavioural treatments of patients with arthritis, fibromyalgia, temporomandibular disorders, musculoskeletal pain and unspecified chronic pain. We found no eligible RCTs on psychodynamic, interpersonal or other psychotherapeutic approaches. The AC received relaxation or education, the WLC contained waiting-list, 'normal activity', 'assessment-only' and medical 'treatment as usual' groups. The characteristics of included RCTs are presented in Supporting Information Tables S3–S5.

Dropout rates and reasons for withdrawal are shown in Supporting Information Fig. S1 and Table S10. Pharmacological studies stated the reasons for withdrawal. Physio- and psychotherapy studies only provided dropout rates in total. Further specifications such as the time of dropout were not given.

### Synthesis of results

At the end of the study, patients treated with drugs showed statistically significant pain reduction relative to their placebo control groups (WMDs) of 8.4 points for WHO I, 10.6 points for WHO II and 12.0 points for WHO III on a 100-point scale (Table 1,  $P < 0.001$ ; Supporting Information Figs. S2–S4). Relative to AC, physiotherapy decreased pain by

4.5 points [not significant (n.s.)] and psychotherapy by 5.5 points (n.s.; Supporting Information Figs. S5 and S6). Compared with WLC these effects were 14.7 points (n.s.) for physiotherapy and 12.1 points for psychotherapy ( $P < 0.01$ ). The corresponding SMDs are presented in Supporting Information Figs. S7–S11. Pre-post comparisons corrected for control group effects yielded similar results (Table 1, right column and Supporting Information Figs. S12–S19). As indicated by confidence intervals (CIs) (Supporting Information Figs. S3 and S4), effect sizes (WMDs, SMDs) did not differ significantly between parallel and crossover drug studies and were therefore pooled regardless of study design. Effect sizes did not differ significantly between drug studies with completely published and those with computed data.

### Comparative consideration of efficacies

Because not enough eligible head-to-head trials between opioids and non-opioids are available (Furlan *et al.*, 2006; Finnerup *et al.*, 2010; Watson *et al.*, 2010; Chaparro *et al.*, 2013), our analysis is limited to adjusted indirect comparisons of outcome data between the five interventions. The heterogeneity of pre-post pain differences in the control groups of drug studies (Supporting Information Figs. S13, S15 and S17) did not allow the definition of a common comparator. Thus, basic requirements for adjusted indirect comparisons between drug categories were not met. The 95% CIs of

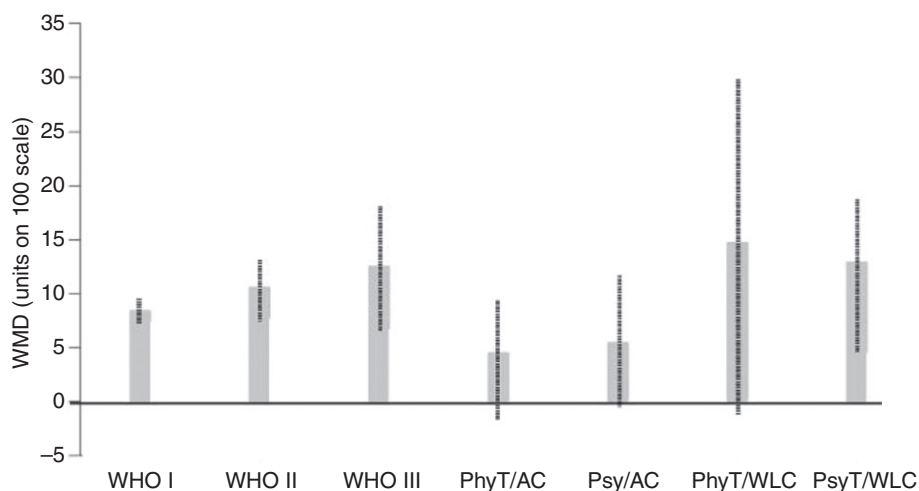
**Table 1**

WMD between pain scores of treatment and control groups at study end (post-post), and between study end (post) and baseline (pre) separated for treatment and control groups<sup>a</sup>

Treatment	WMD post minus post treatment versus control		WMD post minus pre treatments		WMD post minus pre controls		Net effect
	Net effect	(95% CI)	Total effect <sup>b</sup>	(95% CI)	Context effect	(95% CI)	
WHO I – placebo	–8.4	(–9.53, –7.32)	–22.2	(–24.32, –20.11)	–14.9	(–18.49, –11.24)	–7.4
WHO II – placebo	–10.6	(–13.48, –7.64)	–25.4	(–29.12, –21.65)	–14.0	(–19.11, –8.79)	–11.4
WHO III – placebo	–12.0	(–17.98, –6.05)	–29.5	(–37.27, –21.63)	–17.6	(–20.97, –14.13)	–11.9
Physiotherapy – AC	–4.5	(–10.55, 1.50)	–17.8	(–25.14, –10.35)	–12.7	(–20.67, –4.79)	–5.0
Psychotherapy – AC	–5.5	(–11.42, 0.50)	–9.8	(–13.13, –6.54)	–1.3	(–9.22, 6.61)	–8.5
Physiotherapy – WLC	–14.7	(–30.56, 1.15)	–19.0	(–28.09, –9.90)	–3.9	(–11.19, 3.47)	–15.1
Psychotherapy – WLC	–12.1	(–19.65, –4.54)	–13.3	(–17.44, –9.08)	0.2	(–3.55, 3.91)	–13.1

<sup>a</sup>CI of pre-post WMD and net effects are based on SDs of means, not of differences.

<sup>b</sup>Combination of net and context effects in treatment groups.

**Figure 2**

Analgesic effects achieved at end of treatment. Estimates of achievable analgesic efficacies at end of treatment (WMD between control and treatment groups with 95% CIs) for drugs versus placebo (WHO I, II, III); physiotherapy (PhyT) and psychotherapy (PsyT) versus WLC or AC respectively.

our meta-analyses using the outcomes of control groups may serve as the best achievable common comparator (Figure 2). Because of overlapping, the 95% CIs for drug and non-drug treatments do not indicate statistical differences between efficacies (Edwards *et al.*, 2009).

### Functioning and quality of life

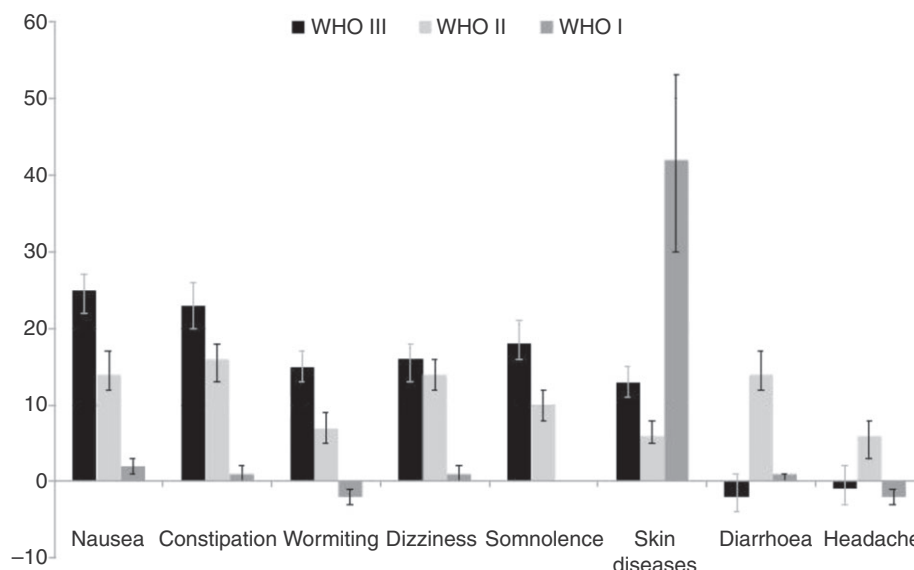
Functioning and quality of life were assessed in 23 of 24 included drug studies, in 8 of 9 included physiotherapy trials, and in 11 of 13 included psychotherapy studies. The results were suitable for descriptive analysis only (Supporting Information Tables S6–S8). Non-opioids (WHO I) and WHO II opioids showed a trend to enhance physical function. Opioids (WHO II and III) tended to improve sleep quality but

did not influence overall health-related quality of life. WHO I studies did not assess the latter aspects. Physiotherapy and psychotherapy had no significant effects on health-related quality of life, sleep quality or physical functioning.

### Adverse effects

All except one drug study reported adverse effects separately for treatment and placebo groups. Risk differences were calculated for each of these effects (Figure 3). Their incidence was 10–20% higher in opioid than placebo groups in the WHO II and III studies. Insufficient data were given on tolerance and addiction. In WHO I trials adverse events were rarely reported, except for one record of skin reactions. No adverse effects were reported in non-pharmacological studies.





**Figure 3**

Common adverse events in drug studies. Risk differences for adverse events between treatment and placebo groups; 95% CIs.

### Subgroup analyses

Formal subgroup analyses for pain syndromes could not be conducted because the respective numbers of studies were too small. For subgroup analyses concerning (median) trial duration, the studies were differentiated into the categories 3–6 weeks (short) and >6 weeks (long). As only one physiotherapy study lasted less than 8 weeks, no subgroup analyses were performed for physiotherapy. Pain reduction as defined by WMDs of pain scores at the end of the study decreased with a longer duration for WHO II drugs (15.4 in six short vs. 7.6 in three long studies), and for psychotherapy compared with AC (7.6 in five short vs. 1.1 in four long studies). These values increased for WHO III drugs (10.2 in four short vs. 15.5 in two long studies), and were similar for WHO I drugs (8.6 in five short vs. 8.6 in four long studies) and for psychotherapy compared with WLC (11.9 in four short vs. 12.8 in three long studies).

### Risk of bias across studies

Comparison of low- with high-quality studies did not yield significantly different values of pain reduction (Supporting Information Figs. S22–S24). Estimated small study effects are shown in Supporting Information Figs. S25–S31. Although the *P*-values of the intercept were below 0.05 for WHO II versus placebo and for psychotherapy versus WLC studies, the adjusted effect sizes (by Duval and Tweedie's trim and fill) did not differ from the calculated effect sizes (Supporting Information Table S9).

## Discussion

We found that the overall analgesic effect of any treatment *per se* was at maximum 15 of 100 points above its control, there was no significant difference between the five treat-

ments, and the net benefit of opioids relative to other treatments was smaller than the efficacy differences between each active treatment and its control. Our analyses are based solely on high-quality RCTs. In such studies (i) pain syndromes are mostly chosen based on previous positive clinical or experimental observations; (ii) patients are carefully pre-selected to minimize variance and to maximize treatment outcomes; and (iii) typically neither authors nor (commercial or public) sponsors leave positive results unpublished (Chou *et al.*, 2009a). Thus, our conclusions are derived from optimized conditions and likely reflect the best possible outcomes regarding pain relief. Even though our analysis was limited to studies published before 2010, a preliminary review of RCTs published thereafter has indicated that no major changes of the present conclusions will be required. The present results extend other meta-analyses in that they enable comparison of average values of pain decrease between opioid, non-opioid and non-pharmacological treatments for CNCP (Chou *et al.*, 2009b; Management of Opioid Therapy for Chronic Pain Working Group, 2010; National Opioid Use Guideline Group, 2010; Chaparro *et al.*, 2013; McNicol *et al.*, 2013).

Some limitations have to be considered. Whereas methodologically adequate RCTs are adequate for causal inferences, meta-analyses are only descriptive and have been criticized for heterogeneity, publication bias and reporting bias. Accepting methodologically poor RCTs for meta-analyses may lead to overemphasis of selected ('positive') results (Chan and Altman, 2005; Kirkham *et al.*, 2010). Studies of low-quality and/or small effect sizes often remain unpublished or are published without statistical parameters which may disclose non-significant results (Hopewell *et al.*, 2007; Bourgeois *et al.*, 2010; Kirkham *et al.*, 2010). Taking these considerations into account, our effect sizes are likely higher than those achievable in reality. This is further supported by uncontrolled observations that even patients who responded to opioids while participating in an RCT termi-

nated their opioid intake later, mostly because of adverse effects or insufficient pain relief (Noble *et al.*, 2008; 2010; Trescot *et al.*, 2008). This may be the reason why outcome data of many long-term trials were not published (Noble *et al.*, 2008). Consistent with our results, patients taking opioids for more than a year in observational studies did not report markedly lower pain scores (58 to 62 of 100 points) than those in RCTs (67 of 100) or those undergoing other treatments in specialized pain clinics (72 of 100) (Dillie *et al.*, 2008; Frettlöh *et al.*, 2009; Sullivan *et al.*, 2010). Another limitation is that even RCTs with the same controls (e.g. placebo) and outcome parameters (e.g. pain intensity) may not be directly comparable (Wells *et al.*, 2009). However, they can still provide hints on maximum achievable values of pain reduction. To assess this upper margin of efficacy, we presented the results of all methodologically acceptable RCTs in forest plots (see Supporting Information Figs. S2–S24). Although the heterogeneity of studies, the lack of head-to-head RCTs comparing different treatments, and the poor description of outcome parameters allowed only an estimation, these values may indicate maximum efficacies that are rarely reached in practice.

The question arises whether the effect sizes determined in our analysis are clinically relevant. For several types of CNCP, a 30% reduction of pain scores, corresponding to a WMD of 20 to 22, is considered a minimal important change (Farrar *et al.*, 2001; Ostelo *et al.*, 2008; Chou *et al.*, 2009b; Dworkin *et al.*, 2009; Mease *et al.*, 2011; Stauffer *et al.*, 2011). We found only 2 out of 41 drug versus placebo comparisons with a WMD of pain scores exceeding 20, and both had extremely high dropout rates. These findings are consistent with other meta-analyses on the analgesic efficacy of opioids and NSAIDs in CNCP (Bjorndal *et al.*, 2007; Chaparro *et al.*, 2013; McNicol *et al.*, 2013). Clinical effectiveness has been defined as the product of efficacy, tolerability, utility, cost and speed (Moore *et al.*, 2010). Each individually perceived analgesic effect has to outweigh adverse effects to result in a net improvement. Thus, a rough assessment of effectiveness, including data on functioning, adverse effects and quality of life, indicates that none of the treatments *per se* generate clinically important improvements. However, mean differences can conceal individual improvement and their clinical importance should be determined by a combined evaluation of adverse effects, physical and emotional functioning, patient adherence, and global improvement (Dworkin *et al.*, 2009). A low net effect may only indicate that CNCP patients with clinically important improvements are less likely to be found. As individual differences in treatment efficacies cannot be predicted, current guidelines recommend individual medication to be maintained or halted based on the continuous assessment of success rather than on predefined pharmacological, mechanistic or pathological categories (Chou *et al.*, 2009b; Management of Opioid Therapy for Chronic Pain Working Group, 2010; National Opioid Use Guideline Group, 2010).

In summary, our assessment of maximum efficacy showed no significant differences between opioids and other pharmacological and non-pharmacological treatments. Apparently no treatment alone reduces CNCP by a clinically relevant amount in a population. Thus, even though sponsors and authors have likely identified the optimal scenarios for

improvement of their RCT participants during the last two decades, there is no evidence to support the sole or preferential use of opioids. This is consistent with the long held notion that CNCP comprises not only somatic nociceptive mechanisms but is largely dependent on psychological and social variables such as learning, conditioning, cognition, affect, emotions, social and cultural influences, financial aspects of the health care system, litigation and others (Stein, 1997; 2013). Although there is little doubt that opioids can inhibit nociception, it is highly contentious whether opioids improve or deteriorate the other factors perpetuating CNCP. This realization may help to prevent overestimation of drug efficacy, overuse, misuse or unrealistic expectations. Our empirical data should enable differential evaluation of therapeutic approaches in comparison with alternatives and decisions on the continuation, switching or combination of treatments. An interdisciplinary multimodal combination of approaches may achieve more important net effects resulting from incremental efficacy of pharmacological and non-pharmacological treatments, and even of placebo effects when they are generated by different sources. However, this needs to be examined in future studies.

## Acknowledgements

This work was funded by the Deutsche Gesellschaft zum Studium des Schmerzes (German Society for the Study of Pain). C. S. was also supported by the Deutsche Forschungsgemeinschaft (STE 477/9) and Bundesministerium für Bildung und Forschung (01 EC 1004 C; 0101-31P5783). The sponsors had no role in study design, data collection and analysis, interpretation of data, preparation of the paper, or in the decision to submit the paper for publication. Parts of these data were discussed within a panel of experts gathered by the German Society for the Study of Pain to develop clinical guidelines for the long-term use of opioids (<http://www.uni-duesseldorf.de/awmf/ll/041-003.htm>).

## Conflict of interest

Authors declare that they have no conflicts of interest.

## References

- Alexander SPH, Benson HE, Faccenda E, Pawson AJ, Sharman JL Spedding M *et al.* (2013a). The Concise Guide to PHARMACOLOGY 2013/14: G Protein-Coupled Receptors. *Br J Pharmacol* 170: 1459–1581.
- Alexander SPH, Benson HE, Faccenda E, Pawson AJ, Sharman JL Spedding M *et al.* (2013b). The Concise Guide to PHARMACOLOGY 2013/14: Enzymes. *Br J Pharmacol* 170: 1797–1867.
- Allison PD (2003). Missing data techniques for structural equation modeling. *J Abnorm Psychol* 112: 545–557.
- Avouac J, Gossec L, Dougados M (2007). Efficacy and safety of opioids for osteoarthritis: a meta-analysis of randomized controlled trials. *Osteoarthritis Cartilage* 15: 957–965.

- Benyamin R, Trescot AM, Datta S, Buenaventura R, Adlaka R, Sehgal N *et al.* (2008). Opioid complications and side effects. *Pain Physician* 11 (2 Suppl): S105–S120.
- Bjordal JM, Klovning A, Ljunggren AE, Slordal L (2007). Short-term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: a meta-analysis of randomised placebo-controlled trials. *Eur J Pain* 11: 125–138.
- Bohnert AS, Valenstein M, Bair MJ, Ganoczy D, McCarthy JF, Ilgen MA *et al.* (2011). Association between opioid prescribing patterns and opioid overdose-related deaths. *JAMA* 305: 1315–1321.
- Bourgeois FT, Murthy S, Mandl KD (2010). Outcome reporting among drug trials registered in ClinicalTrials.gov. *Ann Intern Med* 153: 158–166.
- Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D (2006). Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *Eur J Pain* 10: 287–333.
- Chan AW, Altman DG (2005). Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 330: 753.
- Chaparro LE, Furlan AD, Deshpande A, Mailis-Gagnon A, Atlas S, Turk DC (2013). Opioids compared to placebo or other treatments for chronic low-back pain. *Cochrane Database Syst Rev* (8): CD004959.
- Chou R, Carson S (2008) Drug Class Review on Long-Acting Opioid Analgesics: Final Report Update 5. Oregon Health & Science University: Portland, OR.
- Chou R, Ballantyne JC, Fanciullo GJ, Fine PG, Miaskowski C (2009a). Research gaps on use of opioids for chronic noncancer pain: findings from a review of the evidence for an American Pain Society and American Academy of Pain Medicine clinical practice guideline. *J Pain* 10: 147–159.
- Chou R, Fanciullo GJ, Fine PG, Adler JA, Ballantyne JC, Davies P *et al.* (2009b). Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. *J Pain* 10: 113–130.
- Dillie KS, Fleming MF, Mundt MP, French MT (2008). Quality of life associated with daily opioid therapy in a primary care chronic pain sample. *J Am Board Fam Med* 21: 108–117.
- Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P *et al.* (2009). Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. *Pain* 146: 238–244.
- Edwards SJ, Clarke MJ, Wordsworth S, Borriell J (2009). Indirect comparisons of treatments based on systematic reviews of randomised controlled trials. *Int J Clin Pract* 63: 841–854.
- Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM (2001). Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 94: 149–158.
- Finnerup NB, Sindrup SH, Jensen TS (2010). The evidence for pharmacological treatment of neuropathic pain. *Pain* 150: 573–581.
- Frettlöh J, Maier C, Gockel H, Zenz M, Huppe M (2009). [Characterization of chronic pain patients in German pain centers: core data from more than 10 000 patients]. *Schmerz* 23: 576–591.
- Furlan AD, Sandoval JA, Mailis-Gagnon A, Tunks E (2006). Opioids for chronic noncancer pain: a meta-analysis of effectiveness and side effects. *CMAJ* 174: 1589–1594.
- Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N (2006). Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol* 59: 7–10.
- Gatti A, Mammucari M, Sabato E, Masucci L, Sabato AF (2011). Adherence and long-term effect of oxycodone/paracetamol in chronic noncancer pain: a retrospective study. *Adv Ther* 28: 418–426.
- Herr KA, Spratt K, Mobily PR, Richardson G (2004). Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. *Clin J Pain* 20: 207–219.
- Hopewell S, McDonald S, Clarke M, Egger M (2007). Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database Syst Rev* (2): MR000010.
- Idris NR, Robertson C (2009). The effects of imputing the missing standard deviations on the standard error of meta analysis estimates. *Commun Stat Simul Comput* 38: 513–526.
- Iverson C, Christiansen S, Flanagan A (2007) *AMA Manual of Style: A Guide for Authors and Editors*, 10th edn. Oxford University Press: New York, NY.
- Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R *et al.* (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 340: c365.
- Management of Opioid Therapy for Chronic Pain Working Group (2010). VA/DoD clinical practice guideline for management of opioid therapy for chronic pain.
- Martell BA, O'Connor PG, Kerns RD, Becker WC, Morales KH, Kosten TR *et al.* (2007). Systematic review: opioid treatment for chronic back pain: prevalence, efficacy, and association with addiction. *Ann Intern Med* 146: 116–127.
- McNicol ED, Midbari A, Eisenberg E (2013). Opioids for neuropathic pain. *Cochrane Database Syst Rev* (8): CD006146.
- Mease PJ, Spaeth M, Clauw DJ, Arnold LM, Bradley LA, Russell IJ *et al.* (2011). Estimation of minimum clinically important difference for pain in fibromyalgia. *Arthritis Care Res* 63: 821–826.
- Melnikova I (2010). Pain market. *Nat Rev Drug Discov* 9: 589–590.
- Moher D, Liberati A, Tetzlaff J, Altman DG (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 151: 264–269, W264.
- Moore RA, Derry S, McQuay HJ, Straube S, Aldington D, Wiffen P *et al.* (2010). Clinical effectiveness: an approach to clinical trial design more relevant to clinical practice, acknowledging the importance of individual differences. *Pain* 149: 173–176.
- National Opioid Use Guideline Group (2010). Canadian guideline for safe and effective use of opioids for chronic non-cancer pain.
- Noble M, Tregear SJ, Treadwell JR, Schoelles K (2008). Long-term opioid therapy for chronic noncancer pain: a systematic review and meta-analysis of efficacy and safety. *J Pain Symptom Manage* 35: 214–228.
- Noble M, Treadwell JR, Tregear SJ, Coates VH, Wiffen PJ, Akafomo C *et al.* (2010). Long-term opioid management for chronic noncancer pain. *Cochrane Database Syst Rev* (1): CD006605.
- Nuesch E, Trelle S, Reichenbach S, Rutjes AW, Tschannen B, Altman DG *et al.* (2010). Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ* 341: c3515.
- Okie S (2010). A flood of opioids, a rising tide of deaths. *N Engl J Med* 363: 1981–1985.
- Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korf M *et al.* (2008). Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine* 33: 90–94.



- Papaleontiou M, Henderson CR Jr, Turner BJ, Moore AA, Olkhovskaya Y, Amanfo L *et al.* (2010). Outcomes associated with opioid use in the treatment of chronic noncancer pain in older adults: a systematic review and meta-analysis. *J Am Geriatr Soc* 58: 1353–1369.
- Paulozzi L, Baldwin G, Franklin G, Kerlikowske RG, Jones CM, Ghiya N *et al.* (2012). CDC grand rounds: prescription drug overdoses – a U.S. epidemic. *MMWR Morb Mortal Wkly Rep* 61: 10–13.
- Pizzo PA, Clark NM (2012). Alleviating suffering 101 – pain relief in the United States. *N Engl J Med* 366: 197–199.
- Prakash A, Risser RC, Mallinckrodt CH (2008). The impact of analytic method on interpretation of outcomes in longitudinal clinical trials. *Int J Clin Pract* 62: 1147–1158.
- SIGN (Scottish Intercollegiate Guidelines Network) (2011). Methodology checklist 1: systematic reviews and meta-analyses.
- Solomon DH (2010). Strategies for making analgesia safer: the role of comparative effectiveness research. *Arthritis Rheum* 62: 1568–1570.
- Stauffer ME, Taylor SD, Watson DJ, Peloso PM, Morrison A (2011). Definition of nonresponse to analgesic treatment of arthritic pain: an analytical literature review of the smallest detectable difference, the minimal detectable change, and the minimal clinically important difference on the pain visual analog scale. *Int J Inflamm* 2011: 231926.
- Stein C (1997). Opioid treatment of chronic nonmalignant pain. *Anesth Analg* 84: 912–914.
- Stein C (2013). Opioids, sensory systems and chronic pain. *Eur J Pharmacol* 716: 179–187.
- Stein C, Schäfer M, Machelska H (2003). Attacking pain at its source: new perspectives on opioids. *Nat Med* 9: 1003–1008.
- Sullivan MD, Von Korff M, Banta-Green C, Merrill JO, Saunders K (2010). Problems and concerns of patients receiving chronic opioid therapy for chronic non-cancer pain. *Pain* 149: 345–353.
- Toblin RL, Mack KA, Perveen G, Paulozzi LJ (2011). A population-based survey of chronic pain and its treatment with prescription drugs. *Pain* 152: 1249–1255.
- Trelle S, Reichenbach S, Wandel S, Hildebrand P, Tschannen B, Villiger PM *et al.* (2011). Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. *BMJ* 342: c7086.
- Trescot AM, Helm S, Hansen H, Benyamin R, Glaser SE, Adlaka R *et al.* (2008). Opioids in the management of chronic non-cancer pain: an update of American Society of the Interventional Pain Physicians' (ASIPP) Guidelines. *Pain Physician* 11 (2 Suppl): S5–S62.
- Turk DC, Wilson HD, Cahana A (2011). Treatment of chronic non-cancer pain. *Lancet* 377: 2226–2235.
- Von Korff M, Kolodny A, Deyo RA, Chou R (2011). Long-term opioid therapy reconsidered. *Ann Intern Med* 155: 325–328.
- Watson CP, Gilron I, Sawynok J (2010). A qualitative systematic review of head-to-head randomized controlled trials of oral analgesics in neuropathic pain. *Pain Res Manag* 15: 147–157.
- Wells GA, Sultan SA, Chen L, Khan M, Coyle D (2009) Canadian Agency for Drugs and Technologies in Health (ed.). Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. CADTH: Ottawa.
- Wissenschaftlicher Beirat Psychotherapie (2007). Methodenpapier: verfahrensregeln zur beurteilung der wissenschaftlichen anerkennung von methoden und verfahren der psychotherapie.

Woodcock J (2009). A difficult balance – pain management, drug safety, and the FDA. *N Engl J Med* 361: 2105–2107.

Yates SL, Morley S, Eccleston C, de C Williams AC (2005). A scale for rating the quality of psychological trials for pain. *Pain* 117: 314–325.

## Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

<http://dx.doi.org/10.1111/bph.12634>

**Figure S1** Rates and reasons for dropout in active treatment, placebo control (PC), AC and WLC groups.

**Figure S2** WHO I drug studies with high level of evidence. WMDs between pain scores (0–100) in treatment (analgesic) and placebo groups at study end, based on published (p) or imputed (i) data.

**Figure S3** WHO II drug studies with high level of evidence. WMD between pain scores (0–100) in treatment (analgesic) and placebo groups at study end, based on published (p) or imputed (i) data, separated by study design.

**Figure S4** WHO III drug studies with high level of evidence. WMD between pain scores (0–100) in treatment (analgesic) and placebo groups at study end, based on published (p) or imputed (i) data, separated by study design.

**Figure S5** Physiotherapy studies with a high level of evidence. WMD between pain scores (0–100) in treatment and control groups at study end, separated by type of control group (WLC, AC).

**Figure S6** Psychotherapy studies with high level of evidence. WMD between pain scores (0–100) in treatment and control groups at study end, separated by type of control group (WLC, AC).

**Figure S7** WHO I drug studies with high level of evidence: SMDs between pain scores (0–100) in treatment (analgesic) and placebo groups at study end, based on published (Pizzo and Clark, 2012) or imputed (i) data.

**Figure S8** WHO II drug studies with a high level of evidence: SMD between pain scores (0–100) in treatment (analgesic) and placebo groups at study end, based on published (Pizzo and Clark, 2012) or imputed (i) data, separated by study design.

**Figure S9** WHO III drug studies with high level of evidence: SMD between pain scores (0–100) in treatment (analgesic) and placebo groups at study end, based on published (p) or imputed (i) data, separated by study design.

**Figure S10** Physiotherapy studies with high level of evidence: SMD between pain scores (0–100) in treatment and control groups at study end, separated by type of control group (WLC, AC).

**Figure S11** Psychotherapy studies with a high level of evidence: SMD between pain scores (0–100) in treatment and control groups at study end, separated by type of control group (WLC, AC).

**Figure S12** Treatment groups in WHO I drug studies with a high level of evidence: WMD between pain scores (0–100) at baseline (pre) and study end (post), based on published (p) or imputed (i) data.

**Figure S13** Placebo groups in WHO I drug studies with high level of evidence: WMD between pain scores (0–100) at baseline (pre) and study end (post), based on published (p) or imputed (i) data.

**Figure S14** Treatment groups in WHO II drug studies with high level of evidence: WMD between pain scores (0–100) at baseline (pre) and study end (post), based on published (p) or imputed (i) data, separated by study design.

**Figure S15** Placebo groups in WHO II drug studies with high level of evidence: WMD between pain scores (0–100) at baseline (pre) and study end (post), based on published (p) or imputed (i) data, separated by study design.

**Figure S16** Treatment groups in WHO III drug studies with high level of evidence: WMD between pain scores (0–100) at baseline (pre) and study end (post), based on published (p) or imputed (i) data, separated by study design.

**Figure S17** Placebo groups in WHO III drug studies with high level of evidence: WMD between pain scores (0–100) at baseline (pre) and study end (post), based on published (p) or imputed (i) data, separated by study design.

**Figure S18** Treatment groups in physiotherapy studies with high level of evidence: WMD between pain scores (0–100) at baseline (pre) and study end (post), separated by type of control group (WLC, AC).

**Figure S19** Control groups in physiotherapy studies with high level of evidence: WMD between pain scores (0–100) at baseline (pre) and study end (post), separated by type of control group (WLC, AC).

**Figure S20** Treatment groups in psychotherapy studies with high level of evidence: WMD between pain scores (0–100) at baseline (pre) and study end (post), separated by type of control group (WLC, AC).

**Figure S21** Control groups in psychotherapy studies with high level of evidence: WMD between pain scores (0–100) at baseline (pre) and study end (post), separated by type of control group (WLC, AC).

**Figure S22** WHO I drug studies with low level of evidence: WMD between pain scores (0–100) in treatment (analgesic)

and placebo groups at study end, based on published (p) or imputed (i) data.

**Figure S23** WHO III drug studies with low level of evidence: WMD between pain scores (0–100) in treatment (analgesic) and placebo groups at study end, based on published (p) or imputed (i) data.

**Figure S24** Physiotherapy studies with low level of evidence: WMD between pain scores (0–100) in treatment and control groups at study end, separated by control group category (WLC, AC).

**Figure S25** Publication bias in included WHO I drug studies.

**Figure S26** Publication bias in included WHO II drug studies.

**Figure S27** Publication bias in included WHO III drug studies.

**Figure S28** Publication bias in included physiotherapy (vs. WLC) studies.

**Figure S29** Publication bias in included physiotherapy (vs. AC) studies.

**Figure S30** Publication bias in included psychotherapy (vs. WLC) studies.

**Figure S31** Publication bias in included psychotherapy (vs. AC) studies.

**Table S1** Measures of functioning.

**Table S2** Quality assessment.

**Table S3** Characteristics of included drug studies.

**Table S4** Characteristics of included physiotherapy studies.

**Table S5** Characteristics of included psychotherapy studies.

**Table S6** Effects of WHO I–III drugs on functioning and health-related quality of life.

**Table S7** Effects of physiotherapy on functioning and health-related quality of life.

**Table S8** Effects of psychotherapy on functioning and health-related quality of life.

**Table S9** Publication bias of included studies.

**Table S10** Dropout rates of included RCTs separated by treatment category.