Topical herbal therapies for treating osteoarthritis

Melanie Cameron¹ and Sigrun Chrubasik²
¹School of Health and Sport Sciences, University of the Sunshine Coast, Maroochydore DC, Australia
²University of Freiburg, Freiburg, Germany

Abstract

Background—Before extraction and synthetic chemistry were invented, musculoskeletal complaints were treated with preparations from medicinal plants. They were either administered orally or topically. In contrast to the oral medicinal plant products, topicals act in part as counterirritants or are toxic when given orally.

Objectives—To update the previous Cochrane review of herbal therapy for osteoarthritis from 2000 by evaluating the evidence on effectiveness for topical medicinal plant products.

Search methods—Databases for mainstream and complementary medicine were searched using terms to include all forms of arthritis combined with medicinal plant products. We searched electronic databases (Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, EMBASE, AMED, CINAHL, ISI Web of Science, World Health Organization Clinical Trials Registry Platform) to February 2013, unrestricted by language. We also searched the reference lists from retrieved trials.

Selection criteria—Randomised controlled trials of herbal interventions used topically, compared with inert (placebo) or active controls, in people with osteoarthritis were included.

Data collection and analysis—Two review authors independently selected trials for inclusion, assessed the risk of bias of included studies and extracted data.
Main results—Seven studies (seven different medicinal plant interventions; 785 participants) were included. Single studies (five studies, six interventions) and non-comparable studies (two studies, one intervention) precluded pooling of results.

Moderate evidence from a single study of 174 people with hand osteoarthritis indicated that treatment with *Arnica* extract gel probably results in similar benefits as treatment with ibuprofen (non-steroidal anti-inflammatory drug) with a similar number of adverse events. Mean pain in the ibuprofen group was 44.2 points on a 100 point scale; treatment with *Arnica* gel reduced the pain by 4 points after three weeks: mean difference (MD) −3.8 points (95% confidence intervals (CI) −10.1 to 2.5), absolute reduction 4% (10% reduction to 3% increase). Hand function was 7.5 points on a 30 point scale in the ibuprofen-treated group; treatment with *Arnica* gel reduced function by 0.4 points (MD −0.4, 95% CI −1.75 to 0.95), absolute improvement 1% (6% improvement to 3% decline)). Total adverse events were higher in the *Arnica* gel group (13% compared to 8% in the ibuprofen group): relative risk (RR) 1.65 (95% CI 0.72 to 3.76).

Moderate quality evidence from a single trial of 99 people with knee osteoarthritis indicated that compared with placebo, *Capsicum* extract gel probably does not improve pain or knee function, and is commonly associated with treatment-related adverse events including skin irritation and a burning sensation. At four weeks follow-up, mean pain in the placebo group was 46 points on a 100 point scale; treatment with *Capsicum* extract reduced pain by 1 point (MD −1, 95%CI −6.8 to 4.8), absolute reduction of 1%(7%reduction to 5% increase). Mean knee function in the placebo group was 34.8 points on a 96 point scale at four weeks; treatment with *Capsicum* extract improved function by a mean of 2.6 points (MD −2.6, 95% CI −9.5 to 4.2), an absolute improvement of 3% (10% improvement to 4% decline). Adverse event rates were greater in the *Capsicum* extract group (80% compared with 20% in the placebo group, rate ratio 4.12, 95% CI 3.30 to 5.17). The number needed to treat to result in adverse events was 2 (95% CI 1 to 2).

Moderate evidence from a single trial of 220 people with knee osteoarthritis suggested that comfrey extract gel probably improves pain without increasing adverse events. At three weeks, the mean pain in the placebo group was 83.5 points on a 100 point scale. Treatment with comfrey reduced pain by a mean of 41.5 points (MD −41.5, 95% CI −48 to −34), an absolute reduction of 42% (34% to 48% reduction). Function was not reported. Adverse events were similar: 6%(7/110) reported adverse events in the comfrey group compared with 14% (15/110) in the placebo group (RR 0.47, 95% CI 0.20 to 1.10).

Although evidence from a single trial indicated that adhesive patches containing Chinese herbal mixtures FNZG and SJG may improve pain and function, the clinical applicability of these findings are uncertain because participants were only treated and followed up for seven days. We are also uncertain if other topical herbal products (Marhame-Mafasel compress, stinging nettle leaf) improve osteoarthritis symptoms due to the very low quality evidence from single trials.

No serious side effects were reported.

Authors’ conclusions—Although the mechanism of action of the topical medicinal plant products provides a rationale basis for their use in the treatment of osteoarthritis, the quality and quantity of current research studies of effectiveness are insufficient. *Arnica* gel probably improves symptoms as effectively as a gel containing non-steroidal anti-inflammatory drug, but with no better (and possibly worse) adverse event profile. *Comfrey* extract gel probably improves pain.
and Capsicum extract gel probably will not improve pain or function at the doses examined in this review. Further high quality, fully powered studies are required to confirm the trends of effectiveness identified in studies so far.

**Medical Subject Headings (MeSH)**

Arnica; Capsaicin [therapeutic use]; Comfrey [chemistry]; Drugs, Chinese Herbal [administration & dosage]; Hand Joints; Osteoarthritis [*drug therapy*]; Osteoarthritis, Knee [drug therapy]; Phytotherapy [*methods*]; Plant Extracts [*administration & dosage*]

**MeSH check words**

Humans

---

**PLAIN LANGUAGE SUMMARY**

**Topical herbal therapy for treating osteoarthritis**

This summary of a Cochrane review presents what we know from research about the effects of herbal therapies applied to the skin in people with osteoarthritis.

**The review shows that in people with osteoarthritis**

- *Arnica* gel probably improves pain and function as well as non-steroidal anti-inflammatory drugs do;
- *Capsicum* extract gel probably will not improve pain or function more than placebo;
- Comfrey extract gel probably improves pain more than placebo;
- Chinese herbal patches probably improve pain and function slightly more than placebo.

Herbal therapies may cause side effects; however we do not have precise information about side effects and complications. This is particularly true for rare but serious side effects. Possible side effects may include skin irritations.

**What is osteoarthritis and what is herbal therapy?**

Osteoarthritis (OA) is a disease of the joints (commonly knee, hip, hands). When joints lose cartilage, bone grows to try to repair the damage. Instead of making things better, however, the bone grows abnormally and makes things worse. For example, the bone can become misshapen and make the joint painful and limit movement. OA can affect your physical function, particularly your ability to use your joints.

Herbal medicines are defined as finished, labeled medicinal products that contain as active ingredients aerial or underground parts of plants, other plant material, or combinations thereof, whether in the crude state or as plant preparations (for example oils, tinctures).

**Best estimate of what happens to patients with osteoarthritis who apply *Arnica extract gel***

*Arnica* gel was compared to ibuprofen (a non-steroidal anti-inflammatory).
Pain—(higher scores mean more severe pain): people who applied Arnica rated their pain to be 3.8 points lower (10.1 points lower to 2.5 points higher) than people who applied ibuprofen. After 3 weeks of treatment, people who applied Arnica rated their pain to be 40.4 and people who applied ibuprofen rated their pain to be 44.2 on a scale of 0 to 100.

Physical function—(lower scores mean better function): people who applied Arnica rated their physical function to be 0.4 points lower (1.75 points lower to 0.95 points higher) than people who applied ibuprofen. After 3 weeks of treatment, people who applied Arnica rated their physical function to be 7.1 on a scale of 0 to 30, and people who applied ibuprofen rated their physical function to be 7.5.

Side effects—a greater proportion of people who applied Arnica reported side effects than did those who applied ibuprofen. Fourteen out of 105 people reported side effects with Arnica, and 8 out of 99 people reported side effects with ibuprofen.

Best estimate of what happens to patients with osteoarthritis who apply Capsicum extract gel

Capsicum extract gel was compared to placebo.

Pain—(higher scores mean more severe pain): people who applied Capsicum rated their pain to be 1.0 point lower (6.76 points lower to 4.76 points higher) than people who applied placebo. After 4 weeks of treatment, people who applied Capsicum rated their pain to be 44.6, and people who applied placebo rated their pain to be 45.6 on a scale of 0 to 100.

Physical function—(lower scores mean better function): people who applied Capsicum rated their physical function to be 2.64 points lower (9.51 points lower to 4.23 points higher) on a 0 to 96 point scale than people who applied placebo. After 4 weeks of treatment, people who applied Capsicum rated their physical function to be 32.15 on a scale of 0 to 96, and people who applied ibuprofen rated their physical function to be 34.79.

Side effects—more adverse events were reported among people who applied Capsicum than for those who applied placebo. Of the 338 adverse events reported, 272 occurred in people who applied Capsicum and 66 occurred in people who applied placebo.

Best estimate of what happens to patients with osteoarthritis who apply comfrey extract cream

Comfrey extract cream was compared to placebo.

Pain—(higher scores mean more severe pain): people who applied comfrey rated their pain to be 16.3 points lower (20.08 to 12.58 points lower) than people who applied placebo. After 3 weeks of treatment, people who applied comfrey rated their pain to be lower by 20.9 points from baseline, and people who applied placebo rated their pain to be lower by 4.6 points from baseline on a scale of 0 to 100.
Side effects—a smaller proportion of people who applied comfrey reported side effects than did those who applied placebo. Seven out of 110 people reported side effects with comfrey, and 15 out of 110 people reported side effects with placebo.

Chinese herbal medicine patches

Adhesive patches containing the Chinese herbal mixtures FNZG and SJG were compared to placebo. We are uncertain whether Chinese herbal patches affect osteoarthritis because this intervention was tested over seven days only.

Pain—(higher scores mean worse or more severe pain): people who applied FNZG rated their pain to be 1.44 points lower (9.28 points lower to 6.40 points higher) and people who applied SJG rated their pain to be 1.08 points lower (6.28 points lower to 8.40 points higher) than people who applied placebo. People who applied FNZG rated their pain to be lower by 19.20 points from baseline, people who applied SJG rated their pain to be lower by 16.04 points from baseline, and people who applied placebo rated their pain to be lower by 17.68 points from baseline on a scale of 0 to 100.

Physical function—(lower scores mean better function): people who applied FNZG rated their function to be 2.61 points lower (9.50 points lower to 4.28 points higher) and people who applied SJG rated their function to be 2.97 points lower (9.60 points lower to 3.66 points higher) than people who applied placebo. People who applied FNZG rated their physical function to be lower (better) by 5.04 points from baseline, people who applied SJG rated their physical function to be lower (better) by 6.71 points from baseline, and people who applied placebo rated their physical function to be lower (better) by 6.10 points from baseline on a scale of 0 to 96.

Side effects—a greater proportion of people who applied herbal patches reported side effects than did those who applied placebo patches. Five out of 60 people reported side effects with FNZG, 4 out of 60 people reported side effects with SJG, and 0 out of 30 people reported side effects with placebo.

Other topical products

We are uncertain whether other topical herbal products affect osteoarthritis pain and function because the evidence available from these studies was of low to very low quality. FNZG patches were compared head-to-head with SJG patches. Marhame-Mafasel compress was compared to placebo. Stinging nettle leaf was compared with two placebos in two different studies of people with osteoarthritis of the thumb or of the knee.

BACKGROUND

At times where extraction and synthetic chemistry were not yet invented, musculoskeletal complaints were treated all over the world with preparations from medicinal plants. Due to a legal decision in Germany in 1978, the Commission E of the Federal Health Agency re-evaluated the herbal drugs (Blumenthal 1998). Table 1 summarizes the monographs of approved medicinal plant parts and their preparations for topical use in the treatment of osteoarthritis (OA) complaints. In the course of the harmonization within Europe, the
monographs of the European Scientific Cooperative on Phytotherapy (ESCOP) appeared continuously thereafter and were summarized in the second edition and a supplement (ESCOP 2003; ESCOP 2009) (Table 2). Parallel to this, the American Herbal Pharmacopeia (www.herbal-ahp.org) has been publishing comprehensive monographs accompanied by a Therapeutic Compendium since 1996, and the WHO its monographs on selected medicinal plants since 1999 (http://apps.who.int/medicinedocs/en/d/Js2200e/). Although the ESCOP, American and WHO monographs are not official, they provide scientific information on the safety, efficacy and quality of medicinal plants and provide recommendations for their use in clinical practice (for example doses, types of preparation, warnings). In contrast, the European Medicines Agency monographs (EMA monographs) serve as a guidance for application dossiers to obtain marketing authorization by the regulatory authorities of the individual countries in the European Union.

In the previous Cochrane review on herbal medicines for OA (Little 2000), oral and topical herbal medicines were considered together. However, due to the fact that the mechanism of action of topical medicinal plant products is different from that of oral products, in that they act as counterirritants via the skin or because they are toxic when orally applied, a separation of topical and oral medicinal plant preparations seemed advisable. For example, nettle leaf is covered with needle-like hairs that on contact pierce the skin injecting irritant substances like formic acid, acetic acid, serotonin or 5-hydroxytryptamine, histamine and acetylcholine (Anonymous 1998), which cause an irritant skin reaction. Already in the middle ages urtication (beating with nettle) belonged to the armentarium of treatments for (osteo)arthritic pain.

Menthol, contained in peppermint or other mint oils, is a topical counterirritant (Yosipovitch 1996). The terpene increases the perception of cooling and attenuates the perception of moderate warming (Green 1992) by triggering the cold-sensitive Transient Receptor Potential Melastatine 8 (TRPM8) receptors in skin sensory neurons (Yudin 2012). TRP-Ankyrin1 (A1), another coldsensing channel, is also involved in the menthol cooling sensation (Karashima 2007). The activation of TRPM8 mediates the menthol spasmyolytic effect (Johnson 2009). In vitro studies demonstrated menthol inhibition of the arachidonic acid cascade (cyclooxygenase-2 (COX-2), lipooxygenase) and cytokine release (Juergens 1998). Local anaesthetic (Galeotti 2001), antioxidative (Ka 2005) and analgesic (Taniguchi 1994) actions are other targets of the menthol mechanism of action, the latter based on a weak kappa opioid receptor agonistic effect (Galeotti 2002) and cumulative inactivation of voltage-gated sodium channels (Gaudioso 2012).

The capsaicinoids, the active principle of Capsicum species, act via the heat-sensitive Transient Receptor Potential Vanilloid-1 (TRPV1) receptors (Hayes 2000). Binding of capsaicin to this target is accompanied by a decrease in membrane resistance, depolarization and activation of synaptosomal neurotransmitter release (Buck&Burks 1986;Huang 2008; Sauer 2001; Zhao 1992). Following the initial activation, which is often associated with heat sensation, desensitization and depletion of neurotransmitters produce the capsaicinoid (expressed as capsaicin) analgesic effect. If capsaicin exposure persists, nerve terminals will degenerate (defunctionalization) (Dedov 2000; Dedov 2001; Nolano 1999), which causes the prolonged analgesic effect after the end of treatment. Other capsaicin effects include the
inhibition of inducible COX-2 mRNA expression (Kim 2003) and LOX (Flynn 1986) and a free radical scavenging activity (Galano 2012; Luqman 2006).

Arnica and comfrey do not act as counterirritants. However, both are for topical use due to systemic toxicity (ESCOP 2003; ESCOP 2009) and should only be applied to intact skin. Arnica and comfrey inhibit COX-1 and COX-2 and have an antioxidative potential (ESCOP 2009; Schröder 1990; Verma 2010). So far, inhibition of LOX (Tornhamre 2001), elastase (Siedle 2002; Siedle 2003), cytokines (Jäger 2009; Klaas 2002; Lyss 1997), transcription factor NF-kappaB (Ekenäs 2008) and AP1 (Jäger 2009) has, however, only been demonstrated for the Arnica species. Some effects seem to be likely for comfrey, for example elastase inhibition (Melzig 2005), based on the comfrey ingredient rosmarinic acid for which inhibition of cytokines (Lee 2006) and anti-inflammatory activity has been demonstrated in various animal experiments (Englberger 1988; Moon 2010).

Description of the condition

Osteoarthritis (OA) is characterized by degeneration of the joints, for example the hip, knee and hand. The condition is widespread. Lawrence and co-workers (Lawrence 2008) estimated that among US adults, nearly 27 million had clinical osteoarthritis in 2005 (up from the estimate of 21 million for 1995). Women are more often affected with OA than men, and prevalence increases with increasing age. Overweight and heavy physical work may explain OA in some cases, but non-mechanical factors and genetic disposition are involved as well (van den Berg 2011; Zhang 2010). Diagnostically, primary OA is distinguished from secondary OA induced by traumatic events and endocrine or metabolic disorders. Both primary and secondary forms result in impaired quality of life due to pain and physical disability (Schmritz 2010). The OMERACT-Osteoarthritis Research Society International (OARSI) response criteria combine pain and functional impairments in the identification of treatment response (Pham 2003; Pham 2004) but unfortunately response criteria are not universally considered in clinical studies, making efficacy comparisons difficult.

Description of the intervention

For the purpose of this review we have adopted the World Health Organization (WHO) guidelines for the definition of medicinal plant products, that is, “…finished, labeled, medicinal products that contain as active ingredients, aerial or underground parts of plants, or other plant material, or combinations thereof, whether in the crude state or as plant preparations. Plant preparations include comminuted or powdered plant materials, extracts, tinctures, fatty or essential oils, and any other substances of this nature. Herbal medicines may contain excipients in addition to the active ingredients. Medicines containing plant material combined with chemically defined active substances, including chemically defined, isolated constituents of plants, are not considered to be herbal medicines.” The WHO also notes that “exceptionally, in some countries herbal medicines may also contain, by tradition, natural organic or inorganic active ingredients which are not of plant origin”, however in this review we have applied the strict definition and excluded herbal products combined with non-herbal materials. (apps.who.int/medicinedocs/en/d/Jh2984e).
How the intervention might work

Medicinal plant products used topically for the treatment of OA act as skin irritants (for example Capsicum extract, stinging nettle leaf) and may also act via the same pathways known for oral medicinal plant products, including inhibition of cyclo-oxygenase-1 and 2 (COX-1, COX-2), lipoxygenase (LOX), pro-inflammatory cytokines and enzymes that participate in cartilage destruction, such as elastase and hyaluronidase (for example Capsicum, Arnica, comfrey extracts) (Cameron 2009). Some broad mechanisms of action have been demonstrated in experimental studies (see Background) but the mechanisms have not yet been elucidated in full detail.

Why it is important to do this review

Topical medicinal plant preparations are part of the armamentarium of traditional treatments used by patients suffering from rheumatic pain conditions. The effectiveness of some medicinal plant products is unknown or unclear, and may be associated with risks of harm. This review is important to summarize the evidence of effectiveness of medicinal plant products used topically for OA, and to update the information on these products that is currently captured in the monographs (see Table 1; Table 2). We have undertaken this research to investigate the effectiveness and adverse side effects of these products in the hope that patients with OA and their healthcare providers may make more informed decisions about the usefulness of these interventions.

In the previous Cochrane review on herbal medicines for OA (Little 2000), oral and topical herbal medicines were considered together. When the update of this review became particularly large, a separation of topical and oral medicinal plant products seemed advisable because a) only oral products are purported to have any effect on joint structure, b) topical herbal medicines may act as counterirritants via the skin (for example nettle, peppermint, Capsicum), and c) some products cannot be administered orally due to systemic toxicity (Arnica, comfrey).

OBJECTIVES

To update the existing Cochrane systematic review (Little 2000) by evaluating the evidence of effectiveness for topical medicinal plant products for the treatment of osteoarthritis (OA) by adding data from relevant randomised controlled trials published in the period from January 2000 to February 2013.

METHODS

Criteria for considering studies for this review

Types of studies—All randomised, controlled (placebo or active control), parallel and crossover trials examining the effects of topical herbal interventions for treating OA.

Types of participants—All persons diagnosed with OA according to the American College of Rheumatology (ACR) criteria (Altman 1986; Altman 1990; Altman 1991) or the equivalent European League Against Rheumatism (EULAR) criteria (Zhang 2009; Zhang 2010a). Studies with samples defined according to vague descriptions (for example 'joint
pain’) were not considered. Studies with participant samples defined according to incomplete or partial ACR or EULAR criteria were included, and notes were provided to identify possible weaknesses in sample selection in these studies.

**Types of interventions**—Any topically applied herbal intervention compared with an inert (placebo) or active control was included. Herbal intervention included any plant preparation (whole, powder, extract, standardised mixture) but excluded homeopathy, aromatherapy, or any preparation of synthetic origin.

In the methods published for the original review (Little 2000) herbal therapies used in conjunction with other treatments or combined with a non-herbal substance were also to be included if the effect of the non-herbal intervention was consistent among all groups and quantifiable such that the effect of the herbal intervention could be determined. In this review, however, we have confined interventions to those that comply with the WHO definition of herbal (http://apps.who.int/medicinedocs/en/d/Jh2984e/1.html). According to WHO, herbal therapy combined with a non-herbal substance is no longer herbal treatment. This definition is important because non-herbal substances may interact with the active principle (sum of action of all ingredients) and change effects, potency and safety profile. Even if the non-herbal substance occurs in the same concentration in the placebo control, as is the case in two excluded studies (Gemmell 2003, McKay 2003), the intervention-control comparison is not valid because the nonherbal substance may enhance the absorption of individual ingredients of the active principle or potentiate or reverse the effect of individual ingredients, thus changing the action of the active principle and not the placebo.

**Types of outcome measures**—The main outcome measures considered were consistent with those used across Cochrane Musculoskeletal Group (CMSG) systematic reviews of interventions for OA: pain, function, adverse events, and quality of life (Altman 1996; Pham 2004). To assess the benefits of treatment:

- pain, measured on a visual analogue scale (VAS) (0 to 100), WOMAC pain subscale (0 to 4 or VAS 0 to 100), numerical rating scale (0 to 3), or other pain scales;
- physical function, measured by a VAS (0 to 100), WOMAC function subscale (0 to 4 or VAS 0 to 100), algofunctional index (0 to 3), or other validated functional scales.

To assess the safety of treatment:

- number of participants reporting any adverse event.

Minor outcomes included:

- general well-being or satisfaction indicator;
- with withdrawals due to adverse events;
- serious adverse events;
- quality of life measured by the Short Form (SF)-36 or other validated scales.
We included the following outcomes in the summary of findings tables (derived from the list of outcomes recommended by the CMSG for inclusion in reviews of interventions for osteoarthritis): pain, function, number of participants experiencing any adverse event, withdrawals due to adverse events, serious adverse events, and quality of life. Because there is no purported mechanism for topical herbal medicines to alter joint structure in OA, we omitted radiographic joint changes as a reported outcome from the summary of findings tables.

**Search methods for identification of studies**

**Electronic searches**—For this review update we searched the following electronic databases from the date of the last search in the previously published version of the review to November 2008, and updated the search again on 21 May 2009, 14 December 2010, 16 May 2011, 30 November 2011, 15 June 2012, and finally on 25 and 27 February 2013.

4. MEDLINE (Ovid MEDLINE® In-Process & Other Non-Indexed Citations, to 25 February 2013).
5. EMBASE (via Ovid 2000 to 2011 Week 47)
6. CINAHL (via Ovid 2000 to 2008 Week 5; via EBSCO Host 2008 to 27 February 2013).
7. AMED (via Ovid, 1985 to 30 November 2011).

Thesaurus and free text searches appropriate to each database were performed to combine terms describing OA and terms describing herbal medicine. No methodological filter was applied and the search was not limited by language.

The full search strategies for each database are outlined in Appendix 1.

**Searching other resources**—We searched reference lists of included trials for any other potential studies.

**Data collection and analysis**

**Selection of studies**—This review was an update of a previous review. Two authors of the original review (CL, TP) and two other colleagues (JG, AB) made some contributions to this review and are acknowledged here as investigators, but because these investigators did
not contribute to the totality of the review they are identified in the Acknowledgements rather than listed as authors of this review.

All titles and abstracts identified from electronic databases and other searches were independently examined by three investigators (MC, SC, CL). The full manuscript was retrieved for each record that had the possibility of meeting the review criteria. Three investigators (MC, SC, CL) independently assessed the eligibility of retrieved studies for the review according to the inclusion criteria.

**Data extraction and management**—Data were extracted from each eligible study by two review authors acting independently. Because of the length of time taken to complete this review, and the associated review of oral medicinal plant products for OA, three investigators (MC, SC, TP) contributed to the data extraction.

Two review authors (MC, SC) independently extracted the following data from the included trials and entered the data into RevMan 5:

1. trial characteristics including size and location of the trial, and source of funding;
2. characteristics of the study population including age, and characteristics of the disease including diagnosis criteria and disease duration;
3. characteristics of the therapy in all trial arms including type and dose of therapy;
4. risk of bias domains as outlined in 'Assessment of risk of bias in included studies', below;
5. outcome measures as mean and standard deviation for continuous outcomes, and number of events for dichotomous outcomes (as outlined in Types of outcome measures).

If data were provided for a trial on more than one pain scale, we referred to a previously described hierarchy of pain-related outcomes (Juni 2006; Reichenbach 2007) and extracted data on the pain scale that was highest on this list:

1. global pain;
2. pain on walking;
3. Western Ontario and McMaster Universities Index of Osteoarthritis (WOMAC) osteoarthritis index pain subscore;
4. composite pain scores other than WOMAC;
5. pain on activities other than walking;
6. rest pain or pain during the night;
7. WOMAC global algofunctional score;
8. Lequesne osteoarthritis index global score;
9. other algofunctional scale;
10. patient’s global assessment;
11. physician’s global assessment.

If data on more than one function scale were provided for a trial, we extracted data according to the hierarchy presented below:

1. global disability score;
2. walking disability;
3. WOMAC disability subscore;
4. composite disability scores other than WOMAC;
5. disability other than walking;
6. WOMAC global scale;
7. Lequesne osteoarthritis index global score;
8. other algofunctional scale;
9. patient’s global assessment;
10. physician’s global assessment.

If pain or function outcomes were reported at several time points, we extracted the measure at the end of the intervention as the main outcome.

If data on more than one quality of life scale were provided for a trial, we extracted data according to the hierarchy presented below:

1. SF-36;
2. EuroQoL;
3. SIP (Sickness Impact Profile);
4. NHP (Nottingham Health Profile).

Adverse events were measured as the number of patients experiencing any adverse event, patients who were withdrawn or dropped out because of adverse events, and patients experiencing any serious adverse events. Serious adverse events were defined as events resulting in in-patient hospitalisation, prolongation of hospitalisation, persistent or significant disability, congenital abnormality or birth defect of offspring, life-threatening events, or death.

If additional data were required, we contacted the trial authors to obtain these data. Some data were converted to normalised scales prior to extraction and reporting. Where data were imputed or calculated (for example standard deviations calculated from standard errors, P values, or confidence intervals; or imputed from graphs or from standard deviations in other trials) we reported these adjustments (see Characteristics of included studies). Any disagreements were resolved by consensus.

**Assessment of risk of bias in included studies**—Two review investigators (MC, SC) independently assessed the risk of bias of each included trial against key criteria:
random sequence generation; allocation concealment; blinding of participants, personnel and outcome assessors; incomplete outcome data; selective outcome reporting; and other sources of bias, in accordance with methods recommended by The Cochrane Collaboration (Higgins 2011). Each of these criteria were explicitly judged as: (a) low, (b) unclear (either lack of information or uncertainty over the potential for bias), or (c) high risk of bias. Potential disagreements were discussed and resolved by referring to the original protocol and, if necessary, arbitration by member(s) of the editorial group.

**Measures of treatment effect**—When possible, the analyses were based on intention-to-treat data (outcomes provided for every randomised participant) from the individual trials. For each trial, we presented outcome data as point estimates with means and standard deviations for continuous outcomes and risk ratios (RRs) with corresponding 95% confidence intervals (CIs) for dichotomous outcomes. Where possible, for continuous outcomes we extracted end of treatment scores rather than change from baseline scores. For continuous data, results were presented as mean differences (MDs) and 95% CIs. We had planned that when different scales were used to measure the same outcome or concept, standardised mean difference (SMD) would be used.

**Unit of analysis issues**—Where a study was defined as a crossover trial, data were extracted only up to the point of crossover given the potential for carry-over effects of these particular interventions to bias the treatment effect following crossover.

**Dealing with missing data**—For dichotomous outcomes, we used the number randomised as the denominator and made the assumption that any participants missing at the end of treatment did not have a positive outcome. For continuous outcomes with no standard deviation reported, if possible we calculated standard deviations from standard errors, P values, or CIs. For one study we converted the VAS data from a 10 cm scale to a 100 mm scale (Kosuwon 2010), and for another study we estimated means and standard deviations from graphical data (Grube 2007). Details of data conversion and imputation are explained in the characteristics of included studies and the associated table (see table Characteristics of included studies).

**Assessment of heterogeneity**—We assessed included trials for clinical homogeneity in terms of participants, interventions and comparators. For studies judged as clinically homogeneous, we quantified the possible magnitude of inconsistency (that is heterogeneity) across studies using the I² statistic, with a rough guide for interpretation as follows: 0% to 40% might not be important; 30% to 60% may represent moderate heterogeneity; 50% to 90% may represent substantial heterogeneity; 75% to 100% considerable heterogeneity (Deeks 2011).

**Assessment of reporting biases**—To examine the possibility of publication bias, we planned to construct funnel plots if at least 10 studies were available for the meta analysis of a primary outcome, however we identified too few trials for this analysis.
We planned to assess the presence of small study bias in the overall meta-analysis by checking if the random-effects model estimate of the intervention effect was more beneficial than the fixed-effect model estimate, but again there were too few trials for this analysis.

**Data synthesis**—As far as data extraction was possible, descriptive results are reported for all included studies. No studies could be subject to meta-analysis.

**Subgroup analysis and investigation of heterogeneity**—Our original plan, in order to explain the heterogeneity between the results of the included studies, was to include subgroup analyses by type and length of intervention. Once the review was divided into two reviews, covering topical and oral interventions separately, there were insufficient data in the trials of topical interventions to justify subgroup analyses according to time of intervention.

**Sensitivity analysis**—We planned a sensitivity analysis to investigate the robustness of the treatment effect on pain and function relative to allocation concealment and participant blinding by removing the trials that reported inadequate or unclear allocation concealment and lack of participant blinding from the meta-analysis to see if this changed the overall treatment effect. There were insufficient data to perform these analyses.

**Summary of findings**—See: Summary of findings for the main comparison; Summary of findings 2; Summary of findings 3.

The main results (pain, function, adverse events, withdrawals due to adverse events, serious adverse events, quality of life) of the review are presented in summary of findings tables (Schunemann 2011a). The overall grading of the evidence using the GRADE approach to classify the evidence for each herbal intervention, as: (a) high, (b) moderate, (c) low, or (d) very low, is included as an indication of confidence in the results of the studies. Effect sizes were reported as relative risk and as number needed to treat (Schunemann 2011b).

**RESULTS**

**Description of studies**

See: Characteristics of included studies.

See: Characteristics of excluded studies.

A total of seven new studies were identified for inclusion in this updated review (Grube 2007; Kosuwon 2010; Randall 2000; Randall 2008; Soltanian 2010; Wang 2012; Widrig 2007). The one study of topical capsaicin that was included in the original review (Deal 1991) was excluded from this review when additional information that was provided by the manufacturer allowed us to identify that the capsaicin was extracted (that is a single extracted ingredient, which is not herbal) and not an extract from *Capsicum* fruits. The term capsaicin may be used to refer to capsaicinoids (extract, expressed as capsaicin) or synthetic or extracted capsaicin (single ingredient).
Two studies were of parallel design, with two groups comparing a herbal intervention to a placebo (inert) control (Grube 2007; Randall 2008). One study compared a herbal product with an active control (Widrig 2007). Another study compared two herbal products against each other as well as against a placebo control in a three-arm trial (Wang 2012). Three studies used crossover designs (Kosuwon 2010; Randall 2000; Soltanian 2010). Four studies were of a confirmatory design, with sufficient statistical power (80%) to identify significant effects at the alpha level 0.05 (Grube 2007; Kosuwon 2010; Wang 2012; Widrig 2007). The other three studies were exploratory, showing trends of effectiveness only. The inclusion of three studies is open to question because: (a) participants entered the study with a presumptive diagnosis, not confirmed at baseline, or (b) the criteria by which OA was established were incomplete or inconsistent with ACR or EULAR requirements (Grube 2007; Randall 2000; Randall 2008).

Results of the search—This review was formed from the division of a broader review of herbal therapies for the treatment of OA. In the original review both topical and oral medicinal plant products were considered. The search strategy for this updated review was structured from the protocol used in the original review. The searches for this review update have been repeated several times since 2005. The most recent full search (December 2011) was completed before the current review was divided into two parts. Therefore, it is not possible to give an entirely accurate presentation of the search results as the number of references identified from the search. In the most recent full search of all databases we identified, after the removal of duplicates, 288 abstracts on topical or oral herbal medicines in the treatment of OA. From these abstracts, we identified only one new study that fulfilled the inclusion criteria for this divided review of topical medicinal plant products only. In more recent repeat searches (June 2012 and February 2013) we identified 1771 abstracts, reduced to 159 abstracts after removal of duplicates from previous searches, and from these abstracts four new studies were identified: one that fulfilled the criteria for inclusion, one that was excluded, and two studies available only in abstract form that are currently awaiting classification. See Figure 1 for our best estimate of results from the searches.

Included studies—See: Characteristics of included studies.

Medicinal plant products used for the treatment of OA included crude stinging nettle leaf, standardised extracts from single plants (Arnica, Capsicum, and comfrey), and three mixtures of preparations from multiple plants known as Marhame-Mafasel, Fufang Nanxing Zhitong Gao (FNZG), and Shangshi Jietong Gao (SJG) (proprietary names) (see Table 3 for preparation details of all products).

A few key outcome measures were used but the reporting of measures differed among studies limiting the utility of studies for metaanalysis. All VAS were 100mm lines, with anchor points identified as 0 (nil symptom) and 100 (worst possible symptom), but some authors reported VAS scores on a centimetre scale in the range 0 to 10. For ease of comparison between trials, we converted all VAS data to the 0 to 100 mm scale.

Several studies used WOMAC, but this index may be used with two possible scoring methods: a battery of 0 to 4 Likert scales or a battery of 100 mm VAS. Typically, the Likert
scale scores are presented as aggregate scores (sums) for each of the three subscales (pain subscore range 0 to 20, stiffness subscore range 0 to 8, physical function subscore range 0 to 68), whereas the VAS are converted to normalised units (means) for each subscale (all subscales scored 0 to 100). Although both scoring systems are acceptable for clinical and research use, there is no agreed conversion ratio between them so studies using differing systems are not comparable. Specific details of all data conversions are included in the Characteristics of included studies.

**Excluded studies**—See: Characteristics of excluded studies.

Reasons for excluding studies were: (a) not a randomised controlled trial (Rayburn 2009; Sagar 1988; Saley 1987; Yuelong 2011), (b) review or discussion paper (Kielczynski 1997; Linsheng 1997; Long 2001), or (c) not a herbal intervention (Altman 1994; Gemmell 2003; McCarthy 1992; McCleane 2000; McKay 2003; Schnitzer 1994; Smith 2011).

**Risk of bias in included studies**

See: Characteristics of included studies, ‘Risk of bias’ tables. The methodological quality of each study was assessed independently by two review authors according to the criteria described in the methods (Higgins 2011; Schunemann 2011a). The quality of the included studies was variable and should be taken into account when interpreting the results. See Figure 2 for a summary of the risk of bias assessment.

Only one study adequately met all six validity criteria (Widrig 2007) and was classified as having low risk of bias. All studies were described as randomised. The method of randomisation was not reported in five studies (Grube 2007; Kosuwon 2010; Randall 2000; Wang 2012; Widrig 2007) but two of these studies were conducted in Germany and reported compliance with the International Harmonisation Conference Good Clinical Practice (IHC GCP) guidelines, which is anchored in German law and requires that adequate randomisation, allocation concealment and blinding were undertaken. Risk of bias in these two studies was assessed as low for these criteria (Grube 2007; Widrig 2007).

**Allocation**—Selection bias was rated as low in studies that recruited patients with diagnoses of OA confirmed according to ACR or EULAR criteria (Altman 1986; Altman 1990; Altman 1991; Zhang 2009; Zhang 2010a). In some studies, diagnostic criteria applied at recruitment were not labelled as ACR or EULAR criteria but were described in sufficient detail to be confident that they were fully consistent with the recommendations of these authorities or they were endorsed by other authorities (for example Chinese Orthopaedic Association criteria) (Wang 2012).

In two studies, ACR or EULAR criteria were not fully considered and these studies have been downgraded to unclear risk of selection bias (Grube 2007; Randall 2008). In one study, selection criteria were so broad as to almost certainly have included recruitment of participants with conditions other than OA (Randall 2000). This study has been classified as having high risk of bias.
Allocation concealment was poorly described in most studies. Allocation concealment was assessed according to the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). We attributed low risk of bias to one study (Randall 2008) in which allocation concealment was inferred from the description of the methods and the two studies in which it could reasonably be expected through reported compliance with ICH GCP guidelines (see Other potential sources of bias) (Grube 2007; Widrig 2007). Allocation concealment could not be determined in any other study; neither could failure to conceal allocation be determined. These studies have been classified as having unclear risk of bias for this criterion.

**Blinding**—Low risk of bias has been attributed to four studies (Grube 2007; Kosuwon 2010; Wang 2012; Widrig 2007) in which the herbal products and placebo or active controls could not be distinguished by colour, size, smell, shape, packaging or treatment regimen. In some studies, descriptions of blinding were not explicit but reference was made to compliance with relevant legislation that mandates blinding (see Other potential sources of bias), therefore we acknowledged that these studies also had low risk of bias.

In one study (Soltanian 2010), the method of blinding was inadequately described and no reference was made to governing guidelines. This study was classified as having an unclear risk of performance and detection bias. Two studies of stinging nettle were judged as having unclear risk despite reporting a complete description of the double-blinding method because we considered that placebo validity and blinding may be compromised by stinging side effects of this intervention (Randall 2000; Randall 2008). Although we considered it highly likely that these studies were sufficiently blinded, we have judged the risk of blinding as unclear. Risk of bias has been judged as high in studies that were open label, single blinded, or where interventions could be clearly distinguished.

In some studies where allocation concealment was inadequately described (see Allocation (selection bias)), it was unclear whether clinical examiners were blinded to treatment (detection bias). We have classified these studies as having unclear risk of bias in blinding of the outcome assessor.

**Incomplete outcome data**—Low risk of bias has been attributed to three studies in which participant withdrawals were fully reported and analyses conducted according to an intention-to-treat model (Grube 2007; Wang 2012; Widrig 2007). In these studies the methods for replacing missing data were fully reported. Unclear risk of attrition bias has been attributed to three studies in which withdrawals were reported but not considered in the anyalyses (per protocol analysis only) (Kosuwon 2010; Randall 2000; Randall 2008). One study reported no participant withdrawals and no missing data (Soltanian 2010) and has been classified as having a low risk of bias for this criterion because in this case a per protocol analysis and intention to-treat analysis should be identical. Studies that neither reported participant withdrawals nor applied any method for replacement of missing data were ascribed as at high risk of attrition bias.

**Selective reporting**—Low risk of bias has been attributed to three studies that use a confirmatory design; reported statistical power, effect, and sample size calculations; and
provided results data in sufficient detail to allow extraction for re-analysis (Kosuwon 2010; Wang 2012; Widrig 2007). We have downgraded to unclear risk of reporting bias three studies that used either exploratory designs with small sample sizes (underpowered) (Randall 2008; Soltanian 2010) or where some data were insufficiently reported to allow extraction for re-analysis (Grube 2007). Examples of selective reporting include providing mean scores only (omission of standard deviations) at some or all time points. Similarly, data reported only as group change scores, percentages, or raw scores without measures of data spread, and data presented in graphical form only, were inadequate for reanalysis. One study was particularly poorly reported and has been classified as having high risk of bias for reporting (Randall 2000).

Other potential sources of bias—Selection bias due to diagnostic criteria (see Allocation (selection bias)) is reported under the heading 'other bias' in the risk of bias tables.

We attributed low risk of bias to studies that recruited and assessed participants consistent with the ACR or EULAR criteria, obtained ethics committee approval, with clinical trials registration, used validated outcome measures, and reported compliance with the Declaration of Helsinki and ICH GCP guidelines. Further, we considered that risk of bias could be assumed to be low if satisfying one of these conditions implied satisfaction of another. For example, the ICH GCP guidelines were recommended in Germany, France, Great Britain and Scandanavia from 1986 onwards, therefore we have assumed that Human Research Ethics Committee approvals granted for studies after this time in these countries necessitated compliance with these guidelines. In 1989, these guidelines were recommended across the European Community (EC) as then constituted. Again, we have assumed that from this date studies conducted in EC countries with ethics committee approval have complied with the guidelines regarding randomisation, allocation concealment, and blinding of participants and assessors. In 1996, compliance with ICH CGP guidelines was required under German law governing clinical trials. The ICH GCP guidelines are now adopted by the WHO and most countries, including many developing countries, are listed as following these guidelines. Formally constituted Human Research Ethics Committees are charged with ensuring that clinical trials are conducted in compliance with these guidelines and associated regional legislation. We have classified as low risk all studies that reported either compliance with ICH GCP guidelines or ethics committee approval, or both (Grube 2007; Kosuwon 2010; Randall 2008; Soltanian 2010; Wang 2012; Widrig 2007). High risk of bias has been attributed to the one study that did not report any form of ethical oversight of compliance with research design guidelines (ICH GCP guidelines or Delaration of Helsinki) (Randall 2000).

Effects of interventions

See: Summary of findings for the main comparison Arnica versus ibuprofen for osteoarthritis of the hand; Summary of findings 2 Capsicum for osteoarthritis of the knee; Summary of findings 3 Comfrey for osteoarthritis of the knee
See: Characteristics of included studies; Additional tables Table 3: Herbal medicinal products used for the treatment of OA.

Single source medicinal plant therapies investigated in studies of confirmatory study design were *Capsicum* (Kosuwon 2010), comfrey (Grube 2007), and *Arnica* (Widrig 2007). Results in two studies favoured the herbal interventions over placebo. The other study was a head-to-head comparison of a herbal intervention with an active control. Two studies of exploratory design investigated topical stinging nettle (Randall 2000; Randall 2008). These studies were conducted by the same team of researchers and reported results favouring the intervention, but only one study included sufficient numerical data suitable for extraction. Because of the stinging sensation produced by this intervention, neither study achieved adequate blinding. The single study of Marhame-Mafasel did not include complete details of the herbal product sufficient to replicate the study (Soltanian 2010). The same was true for the study of Chinese herbal patches (FNZG and SJG), however these products are proprietary and replication of these studies (multiple comparisons in one report) may be possible if the products were prepared according to manufacturing standards (Wang 2012). Results of all comparisons of interventions against placebo and head-to-head comparisons are reported for interest and completeness. No serious side effects were observed with any topical medicinal plant product.

**Arnica montana** (*Arnica*)—Three times daily topical application of a gel containing a tincture of *Arnica montana* was compared with a gel containing ibuprofen in 204 patients (174 participants per protocol) with OA of the hands over three weeks (Widrig 2007). Hand pain measured using a 100 mm VAS, hand function, 28 tender joint count, and duration and intensity of morning stiffness were not significantly different between groups, either as final end point measures or as changes from baseline scores. Mean cumulative doses of rescue medication (acetaminophen) differed only by 25 mg (MD 25, 95% CI 1066.47 to 1016.47; Analysis 1.6) over the intervention period. The number of participants reporting adverse events was similarly consistent between the two groups (odds ratio (OR) 1.75, 95% CI 0.70 to 4.37, P = 0.23; Analysis 1.7). These results suggested that short term topical use of *Arnica* gel afforded not inferior effects to those of ibuprofen gel, consistent with the research hypothesis. No comparison of *Arnica* gel to placebo was identified in this systematic review of the literature.

**Capsicum species**—In 99 patients studied over four weeks, three times daily application of a gel containing a tincture of *Capsicum* species was superior to placebo in reducing osteoarthritic knee pain measured using a 100 mm VAS (MD −1.00, 95% CI −6.76 to 4.76; Analysis 2.1) and overall OA measured using the composite WOMAC score (MD −2.64, 95% CI −9.51 to 4.23; Analysis 2.2). On both these measures the effect sizes were small and CIs crossed the midline indicating that *Capsicum* was not markedly better than placebo. Fifty-seven participants reported a burning sensation in the skin during treatment with *Capsicum* extract gel but no participants with drew from the study for this reason. Burning is a known side effect of *Capsicum*, associated with the mechanism of action of this medicinal plant, and may not be sufficiently problematic to be classified as ‘adverse’, however when
burning was included as an adverse event, the risk ratio of experiencing an adverse event while using *Capsicum* gel rather than placebo was 4.12 (95% CI 3.30 to 5.15; Analysis 2.3).

**Symphytum officinale** (comfrey)—In a large (n = 220) parallel group trial, three times daily topical use of an ointment containing comfrey root (*Symphytum offic. radix*) was compared with placebo over three weeks of intervention (Grube 2007). Grube 2007 found that treatment with comfrey root resulted in statistically significant improvements on the 100 mm VAS measures of total pain, pain at rest, and pain on movement; and on WOMAC scores of pain, stiffness, physical function and overall score. Data from this study could not be extracted for further analysis because the trial authors reported neither absolute scores nor measures of data spread (standard deviations, CIs) for any outcomes (Grube 2007). Mean within-group changes from baseline in pain at rest, pain on movement, WOMAC pain, stiffness, physical function and total scores, and SF-36 physical and mental component summary scores, are reported here for descriptive comparison (see Analysis 3.3 to Analysis 3.10).

**Urtica dioica** (stinging nettle)—Seven days of topical application of one stinging nettle leaf (freshly cut once a day and then applied directly to the painful area with gentle pressure and leaf movement) was compared with placebo (white dead nettle) for base of thumb pain (Randall 2000). This study was of limited use because the diagnosis of OA, although likely, was not established at baseline using ACR or EULAR criteria. This study was a crossover trial with two single weeks of intervention, each preceded by five weeks of washout. Randall 2000 reported that one week of treatment with stinging nettle afforded statistically significant improvements in pain measured using a 100 mm VAS (P = 0.026) and disability measured using the Stanford Health Assessment Questionnaire Disability Index (HAQ-DI, P = 0.003) over placebo. Data reported in this study were presented per intervention rather than providing divided data for each stage of the crossover, and were insufficient to allow extraction for reanalysis.

A follow-up study by the same author team was a one week comparison of stinging nettle leaf against another *Urtica* species. This study included 16 weeks of follow-up. In between-group comparisons for pain at one week post-treatment (Analysis 5.1), and stiffness (Analysis 5.2) and physical function (Analysis 5.3) at four weeks post-treatment, stinging nettle was not significantly different to placebo. Because the stinging nettle group showed a greater within-group improvement in pain at one week post-treatment, the authors argued in favour of this treatment, however we noted that the stinging nettle group commenced the study with a greater mean pain score at baseline, so improvement in this group was not hampered by a floor effect.

**Herbal mixture (Marhame-Mafasel):** A pomade of herbs known as Marhame-Mafasel was compared against placebo in a crossover study of 42 participants with OA of the knee. This study comprised two intervention periods of three weeks each. No washout period occurred between the intervention periods but this weakness in study design was accounted for in this review because we have extracted data from the first intervention period only (up to crossover). These results showed small effects for Marhame-Mafasel over placebo for improvements in pain (Analysis 4.1), physical function (Analysis 4.3), knee stiffness
(Analysis 4.2) and overall disease severity (composite WOMAC score; Analysis 4.4). Although the authors reported a large and statistically significant omnibus effect for treatment (mean effect 3.94, SD 2.01), none of the univariate effect sizes appeared to be statistically significant or clinically meaningful (minimal clinically important difference (MCID) not reported). Further, although the authors reported no dropouts or withdrawals from the study, we question the meaningfulness of this claim because compliance with the intervention was low: “A patient was considered to comply with the assigned treatment if more than 75% of the pomade in the tubes was taken and moderate compliance if 25% to 75% of the pomade in the tubes was taken”. Participants who used less than 25% of the pomade were classified as having poorly complied with the intervention, yet data from these participants were included unaltered in the study. It was possible that this classification of compliance was created post hoc as a strategy to include all data. We suggest that monitoring throughout the study and exclusion of non-complying participants, with replacement of missing data via the last observation carried forward method, would have been more robust and meaningful. Alternately, a post hoc multivariate analysis could have been undertaken to determine any confounding effect of poor participant compliance.

**Chinese herbal patches:** Chinese herbal patches containing either Fufang Nanxing Zhitong Gao (FNZG) or Shangshi Jietong Gao (SJG) were compared to placebo in a three-arm trial of 150 participants with OA of the knee. The intervention was maintained for seven days. The results showed modest effects in favour of both Chinese herbal patches over placebo, with effects being slightly larger in the FNZG group. Although the study was of a confirmatory design with sufficient power (80%) to detect changes, none of the effects were statistically significant. Participants in the FNZG patch group rated their pain on walking (Analysis 6.1), pain due to OA (Analysis 6.2), and physical function (Analysis 6.4) as improved, compared with participants who used the placebo patches, but they also reported more adverse side effects (Analysis 6.6), notably skin irritation. Results were noted in a similar direction but with smaller effect sizes for SJG patches over placebo for pain on walking (Analysis 7.1), pain due to OA (Analysis 7.2), and physical function (Analysis 7.4); as well as similar rates of side effects (Analysis 7.6). A head-to-head comparison of the two patches was equivocal. No participants reported adverse effects from using the placebo patches.

**DISCUSSION**

**Summary of main results**

One confirmatory study is available for products from *Arnica montana* (Widrig 2007) (Summary of findings for the main comparison), *Capsicum* species (Kosuwon 2010) (Summary of findings 2), *Symphytum officinale* (Grube 2007) (Summary of findings 3), and two Chinese herbal patches (Wang 2012) (Table 4; Table 5; Table 6). Moderate quality evidence from one trial (174 participants) indicates that *Arnica montana* is equivalent to topical ibuprofen in terms of pain relief and improvement of hand function. We are less certain about the incidence of adverse events, which may be of concern with both topical
Arnica extract and ibuprofen gel. Moderate evidence from one trial (99 participants) shows that topical Capsicum extract may possibly improve pain and overall function in people with osteoarthritis (OA) of the knee, but improvements are inconsistent (confidence intervals cross the midline) and some people may experience adverse effects, particularly skin irritation and burning.

Moderate evidence from one trial (150 participants) shows that patches containing two different formulations of Chinese herbs may possibly improve pain and function in people with OA of the knee, but the interventions were tested over seven days only, which may be insufficient for making judgements about clinical importance. We are uncertain about the clinical application of this evidence but the trial was quite well designed (double blind, randomised, controlled), thus we have graded the evidence for Chinese herbal patches as moderate but we have presented the summary of findings tables for these interventions under additional tables.

One exploratory study of the herbal mixture Marhame-Mafasel (42 participants) identified a possible trend of effectiveness (confidence interval cross midline) that needs to be investigated in further rigorous trials (Soltanian 2010) (Table 7). Two pilot studies of topical nettle leaf returned disparate results; one study (crossover design) identified a trend for effectiveness (Randall 2000) (Table 8) but the follow-up study (parallel groups) returned equivocal results on between-group comparisons (Randall 2008) (Table 9). Both these studies were hampered by design flaws.

**Overall completeness and applicability of evidence**

The mechanism of action provides a rationale for topical medicinal plant products from Arnica montana, Capsicum species, Symphytum officinale and Urtica dioica as alternative options for the treatment of OA complaints. However, for the herbal mixtures the mechanism of action is less well elucidated through in vitro studies, and the rationale for their use is unclear.

For none of the products is the quality or quantity of current scientific evidence of effectiveness sufficient. There is, at best, moderate evidence to support the use of Arnica, Capsicum and comfrey. However, for each of these interventions, further high quality clinical trials are likely to have an important impact on our confidence in the estimate of effect and may change the estimate. To be more confident in our estimates of clinical effectiveness we require well designed, randomised, double blind studies of a confirmatory study design with adequate power and sample size (n > 400) that test interventions over clinically relevant durations.

The results of studies undertaken with a proprietary product cannot be transferred to any preparation of the medicinal plant part (Chrubasik 2003). If the starting material and manufacturing process of products differ, active principles will differ and thus the sum of all actions of the ingredients. Due to insufficient declaration, the studies undertaken with Arnica, Capsicum, comfrey, and the herbal mixtures FNZG, SJG and Marhame-Mafasel are not repeatable unless the products can be obtained from the producer or the laboratory. Even if these products can be obtained, due consideration must be given to the guidelines of Good
Manufacturing Practice (GMP) and Good Distribution Practice (GDP); these guidelines ensure that medicinal plant products are consistently produced and controlled to the quality standards appropriate to their intended use, and that the level of quality determined by the GMP and the properties of the products are maintained throughout the distribution (www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/document listing/document listing 000154.jsp&mid=WC0b01ac0580027088&jsenabled=tru, www.who.int/vaccines- documents/DocsPDF/www9666.pdf).

It is a common but erroneous assumption that medicinal plant products are safer than other therapies for OA. All topical herbal medicinal products for the treatment of OA, except preparations from *Capsicum* species, have a low risk of adverse events when used in the suggested doses (Table 1; Table 2). Minor adverse reactions occurred with all medicinal plant treatments identified in this review, and only in the case of comfrey were these events more commonly reported among the placebo group (Analysis 3.11). Allergic reactions may occur with any of the topical medicinal plant products (ESCOP 2003; ESCOP 2009), but *Capsicum* species, comfrey and *Arnica* also contain toxic ingredients. Capsaicin is neurotoxic (Anonymous 2007; Nolano 1999) and a potential carcinogen (in animal and in vitro studies) (Anonymous 2007). The alkaloids in comfrey are hepatotoxic and carcinogenic (Li 2011). In vitro studies of *Arnica* raise concerns of cytotoxicity (Woerdenbag 1994). Because of the risk of cytotoxicity, comfrey and *Arnica* are recommended for external use only (ESCOP 2003; ESCOP 2009). In contrast to the other medicinal plant preparations, use of capsaicinoid containing preparations is restricted up to several weeks (ESCOP 2009) and the content of toxic alkaloid in the daily dose of topical comfrey has been limited to 100 µg per day (Blumenthal 1998).

**Quality of the evidence**

See: Characteristics of included studies, ‘Risk of bias’ tables.

Generally, the studies included in this review are of lower quality than desired, but we stress that these studies represent the current best quality evidence for the effectiveness of topical medicinal plant interventions in the treatment of OA.

**Moderate evidence for estimate of effect**—there is, at best, moderate evidence for creams and gels containing *Arnica*, comfrey, or *Capsicum* extract and Chinese herbal patches (FNZG and SJG) as topical herbal medicines in the treatment of OA. The evidence for these interventions is drawn from small (n < 400) single studies and is thus downgraded to moderate. Because the patches containing the two formulations of Chinese herbs were tested over seven days only, which may be insufficient for making judgements about clinical importance, we are uncertain about the clinical application of this evidence. We have graded the evidence for Chinese herbal patches as moderate but have presented the summary of findings table for these interventions under additional tables (Table 4; Table 5; Table 6).

**Low evidence for estimate of effect**—one exploratory study of the herbal mixture Marhame-Mafasel (42 participants) identified a possible trend of effectiveness (confidence intervals cross midline) that needs to be investigated in further rigorous trials (Soltanian 2010) (Table 7).
**Very low evidence for estimate of effect**—two pilot studies of topical nettle leaf returned disparate results; one study (crossover design) identified a trend of effectiveness (Randall 2000) (Table 8) but the follow-up study (parallel groups) returned equivocal results on between-group comparisons (Randall 2008) (Table 9). Both these studies were hampered by design flaws.

Poorer quality studies using non-randomised, uncontrolled designs were excluded (for example Linsheng 1997). Similarly, we excluded clinical trials of products that are not strictly herbal so as to avoid misinterpretation of the results of these studies in herbal medicine practice (for example Altman 1994; Gemmell 2003). We note that more recent studies are typically of higher quality than older studies and commend researchers in this field for the improvement in research design and reporting.

**Potential biases in the review process**

This review is compromised by some poorly designed clinical trials that are underpowered and inadequately blinded. Herbal medicine is not a field known for the widespread adoption of evidence-based practice, however, in light of the small and low quality body of evidence in topical herbal treatment for OA, it is unsurprising that practitioners might continue to ignore the research and do what they 'have always done’. In this section, therefore, we have chosen to address some of the common biases in herbal medicine as well as in this review.

**Agreements and disagreements with other studies or reviews**

Evidence for topical capsaicin in the relief of osteoarthritic pain has previously been described as promising (Cameron 2007; Cameron 2009; Little 2000); however, because extracts reduced to single compounds are not herbal interventions according to the strictest WHO definition, studies investigating the single extracted ingredient capsaicin were excluded from this review. The one study of an extract from *Capsicum* fruits that was included in this review showed small beneficial effects of the intervention, but not significantly greater than with placebo (Analysis 2.1; Analysis 2.2; Summary of findings 2). Favourable effects identified in the excluded studies (Altman 1994; Deal 1991; McCarthy 1992; McClean 2000; Schnitzer 1994) are generally larger but are attributed to higher doses of capsaicin (0.025 to 0.05% v.v. in a vehicle cream) than the dose used in the included study (0.0125%). Even at the lower dose, the extract of *Capsicum* species is associated with a substantive risk of skin irritation (RR 4.12, 95% CI 11.61 to 24.84).

**AUTHORS’ CONCLUSIONS**

**Implications for practice**

The current available evidence for topical herbal treatment of osteoarthritis (OA) is sparse and it is difficult to give clear recommendations regarding use of these products. Generally, high tolerance of the herbal medicinal products was demonstrated; however, caution is warranted in interpreting safety due to the small sample size in some of the studies. Accepting that there are few high quality randomised controlled trials of the efficacy or safety of topical medicinal plant products, in the absence of more robust results we
recommend that practitioners adopt the preparations, methods of administration, and doses
of topical products suggested in the monographs (see Background).

The only recommendations we are confident to make for clinical practice are that a) Arnica
gel probably improves pain and function as effectively as a gel containing non-steroidal anti-
inflammatory drug, but with no better (and possibly worse) adverse event profile; and b)
comfrey extract gel probably improves pain more than placebo. Effects of comfrey gel on
physical function and quality of life in people with OA are not estimable from the data
provided. Capsicum extract gel probably will not improve pain or function more than
placebo at the dose examined in this review. Although patches containing the Chinese herbal
mixtures FNZG and SJG probably slightly improve pain and function more than placebo, we
are uncertain of the clinical applicability of these results because these interventions were
tested over seven days only. There is insufficient evidence to make clinical
recommendations for or against the use of other topical herbal medicines for the treatment of
OA.

Implications for research

We recommend that future updates of this review focus on the topical herbal interventions
for which there currently appears to be moderate evidence, Arnica, Capsicum, comfrey, and
the Chinese herbal mixtures FNZG and SJG.

At this stage we cannot recommend that resources be invested in single small studies of
untested herbal interventions or herbal interventions for which the current evidence is low or
very low. Such studies do not add substantially to the body of evidence but increase
confusion among practitioners.

Several studies were excluded from this review on the grounds that they did not investigate
truly herbal products. Included studies are hampered by flawed research design, including
unclear recruitment criteria, and inadequate characterisation of the herbal interventions.
Other studies are of limited usefulness because the selection criteria were incomplete,
methods were confusingly reported (Begg 1996; Moher 2001), or data were presented to
support the authors’ preferred conclusions (McGauran 2010). We recommend that future
researchers give attention to the detail of study design, ensuring that participant samples are
well defined according to ACR and EULAR criteria and recruited without bias, that herbal
preparations are reported in detail, including dose, extraction method and active principle,
and that study results are recorded using reliable, valid outcome measures.

Evidence for mechanisms of effect and toxicity are drawn from animal studies and in vitro
designs rather than from human clinical trials. Well designed, fully powered clinical trials
are required to confirm the efficacy of most topical medicinal plant products in humans. We
encourage herbal medicine practitioners to consider involvement of themselves, their
practices, and their patients in future clinical trials to ensure that representative patient
groups are included and that trial results have broad applicability to everyday practice.
Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The review authors would like to thank the Cochrane Musculoskeletal editorial team for their editorial suggestions.

Christine Little (CL) and Tessa Parsons (TP) authored the original review that formed the template for this updated version. CL contributed to paper selection for this review, and TP extracted data from some studies. We gratefully acknowledge their contributions to the foundational work for this review.

Rudolf Bauer (RB) and Renea Johnston (RJ) provided advice regarding refinement of the methods for this review. RB edited portions of the Background. RJ provided invaluable support in the presentation of the summary of findings tables and calculations of absolute and relative risk. We thank these colleagues for their support and assistance in finalising this review.

SOURCES OF SUPPORT

Internal sources

- Victoria University, Australia.
  Victoria University provided one author (2004–2009), and allowed time release from normal duties to undertake review training.
- University of Freiburg, Germany.
  University of Freiburg provided one author.
- Australian Catholic University, Australia.
  The Australian Catholic University provided one author with time (2010–2011). Librarians from the Australian Catholic University assisted with the acquisition of full manuscripts of studies included in this review.
- University of the Sunshine Coast, Australia.
  The University of the Sunshine Coast provided one author (2012–2013), and allowed time release from normal duties to complete this review.

External sources

- National Center for Complementary and Alternative Medicine, USA.
  This work was partially funded by Grant Number R24 AT001293 from the National Center for Complementary and Alternative Medicine (NCCAM). The contents of this systematic review are solely the responsibility of the authors and do not necessarily represent the official views of the NCCAM or the National Institutes of Health.

REFERENCES

* Indicates the major publication for the study

References to studies included in this review


References to studies excluded from this review


References to studies awaiting assessment


References to ongoing studies


Additional references


References to other published versions of this review


Little CV, Parsons T, Logan S. Herbal therapy for treating osteoarthritis. Cochrane Database of Systematic Reviews. 2000; (Issue 4)
APPENDICES

Appendix 1. Search Strategies

MEDLINE

1. exp osteoarthritis/
2. osteoarthr$.tw.
3. (degenerative adj2 arthritis).tw.
4. arthrosis.tw.
5. or/1–4
6. exp Medicine, Herbal/
7. exp Plants, Medicinal/
8. exp Medicine, Traditional/
9. exp Drugs, Chinese Herbal/
10. herb$.tw.
11. (plant or plants).tw.
12. phytomedicine.tw.
13. botanical.tw.
14. weed$.tw.
15. algae.tw.
16. (fungi or fungus).tw.
17. ((traditional or chinese or herbal) adj medicine).tw.
18. ((oriental or chinese) adj tradition$).tw.
19. or/6–18
20. 5 and 19

EMBASE

1. exp osteoarthritis/
2. osteoarthr$.tw.
3. (degenerative adj2 arthritis).tw.
4. arthrosis.tw.
5. or/1–4
6. exp Herbal Medicine/
7. exp Medicinal Plant/
8. exp Traditional Medicine/
9. exp Chinese Medicine/
10. herb$.tw.
11. (plant or plants).tw.
12. phytomedicine.tw.
13. botanical.tw.
14. weed$.tw.
15. algae.tw.
16. (fungi or fungus).tw.
17. ((traditional or chinese or herbal) adj medicine).tw.
18. ((oriental or chinese) adj tradition$).tw.
19. or/6–18
20. 5 and 19

**CINAHL**

1. exp OSTEOARTHRITIS/
2. osteoarthr$.tw.
3. (degenerative adj2 arthritis).tw.
4. arthrosis.tw.
5. or/1–4
6. exp Medicine, Herbal/
7. exp Plants, Medicinal/
8. Medicine, Traditional/
9. exp Plant Extracts/
10. herb$.tw.
11. (plant or plants).tw.
12. phytomedicine.tw.
13. botanical.tw.
14. weed$.tw.
15. algae.tw.
16. (fungi or fungus).tw.
17. ((traditional or chinese or herbal) adj medicine).tw.
18. ((oriental or chinese) adj tradition$).tw.
19. or/6–18
20. 5 and 19

Revised Strategy (EBSCOhost)

S24 S5 and S22
S23 S5 and S22
S22 S6 or S7 or S8 or S9 or S10 or S11 or S12 or S13 or S14 or S15 or S16 or S17 or S18 or S19 or S20 or S21
S21 ti chinese tradition* or ab chinese tradition*
S20 ti oriental tradition* or ab oriental tradition*
S19 ti herbal medicine or ab herbal medicine
S18 ti chinese medicine or ab chinese medicine
S17 ti traditional medicine or ab traditional medicine
S16 ti fungi or ti fungus or ab fungi or ab fungus
S15 ti algae or ab algae
S14 ti weed* or ab weed*
S13 ti botanical or ab botanical
S12 ti phytomedicine or ab phytomedicine
S11 ti plant or ti plants or ab plant or ab plants
S10 ti herb* or ab herb*
S9 (MH “Plant Extracts+”)
S8 (MH “Medicine, Traditional+”)
S7 (MH “Plants, Medicinal+”)
S6 (MH “Medicine, Herbal+”)
S5 S1 or S2 or S3 or S4
S4 ti arthrosis or ab arthrosis
S3 ti degenerative N2 arthritis or ab degenerative N2 arthritis

S2 ti osteoarthr* or ab osteoarthr*

S1 (MH “Osteoarthritis+”)

**AMED**

1. exp Osteoarthritis/
2. osteoarthr$.tw.
3. (degenerative adj2 arthritis).tw.
4. arthrosis.tw.
5. or/1–4
6. exp herbal drugs/
7. exp traditional medicine/
8. exp plant extracts/
9. exp plants medicinal/
10. herb$.tw.
11. (plant or plants).tw.
12. phytomedicine.tw.
13. botanical.tw.
14. weed$.tw.
15. algae.tw.
16. (fungi or fungus).tw.
17. ((traditional or chinese or herbal) adj medicine).tw.
18. ((oriental or chinese) adj tradition$).tw.
19. or/6–18
20. 5 and 19

**The Cochrane Library 2008, Issue 4**

#1 MeSH descriptor Osteoarthritis explode all trees

#2 osteoarthr*:ti,ab

#3 (degenerative near/2 arthritis):ti,ab

#4 arthrosis:ti,ab
#5 (#1 OR #2 OR #3 OR #4)

#6 MeSH descriptor Medicine, Herbal explode all trees

#7 MeSH descriptor Plants, Medicinal explode all trees

#8 MeSH descriptor Medicine, Traditional explode all trees

#9 MeSH descriptor Drugs, Chinese Herbal explode all trees

#10 herb*:ti, ab

#11 (plant or plants):ti, ab

#12 phytomedicine:ti,ab

#13 botanical:ti, ab

#14 weed*:ti, ab

#15 algae:ti, ab

#16 (fungi or fungus):ti, ab

#17 ((traditional or chinese or herbal) next medicine):ti, ab

#18 ((oriental or chinese) next tradition*):ti, ab

#19 (#6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18)

#20 (#5 AND #19)

**ISI Web of Science**

#7 #4 AND #1


#6 #4 AND #1

Refined by: Publication Years=(2009 OR 2007 OR 2004 OR 2001 OR 2010 OR 2005

#5 #4 AND #1

#4 #3 OR #2

#3 Topic=((((oriental or chinese or traditional) and (medicine or therap*))))
#2 Topic=(herb* or plant or plants or phytomedicine or botanical or weed* or algae or fungi or fungus)

#1 Topic=(arthrit* or arthrosis or osteoarthrit* or osteoarthrosis)

**Dissertation Abstracts**

arthrit* or arthrosis or osteoarthrit* or osteoarthrosis AND herb* or plant or plants or phytomedicine or botanical or weed* or algae or fungi or fungus or oriental or chinese or traditional in (medicin* or therap*)

**World Health Organization International Clinical Trials Registry Platform**

Osteoarthritis in Condition AND herb* or plant or plants or phytomedicine or botanical or weed* or algae or fungi or fungus or oriental or chinese or traditional in Intervention
Figure 1.
Study flow diagram.
**Figure 2.**
Risk of bias summary: review authors’ judgements about each risk of bias item for each included study.
Analysis 1.1.
Comparison 1 Arnica versus ibuprofen, Outcome 1 Pain VAS 0-100.
Analysis 1.2.
Comparison 1 Arnica versus ibuprofen, Outcome 2 28 painful joint count change from baseline.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Arnica gel</th>
<th>Ibuprofen gel</th>
<th>Mean Difference</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N  Mean(SD)</td>
<td>N  Mean(SD)</td>
<td>IV/Random,95% CI</td>
<td>IV/Random,95% CI</td>
</tr>
<tr>
<td>Widrig 2007</td>
<td>89  -3 (4.4)</td>
<td>85  -2.5 (3.2)</td>
<td></td>
<td>-0.50 [-1.64, 0.64]</td>
</tr>
</tbody>
</table>
### Analysis 1.3.
Comparison 1 Arnica versus ibuprofen, Outcome 3 Intensity of morning stiffness (1 to 5) change from baseline.
Analysis 1.4.
Comparison 1 Arnica versus ibuprofen, Outcome 4 Duration of morning stiffness (1 to 5) change from baseline.
### Analysis 1.5.
Comparison 1 Arnica versus ibuprofen, Outcome 5 Hand algofunctional index (0 to 30).

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Arnica gel</th>
<th>Ibuprofen gel</th>
<th>Mean Difference</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Widrig 2007</td>
<td>N = 89</td>
<td>N = 85</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean(SD) = 7.1 (4.8)</td>
<td>Mean(SD) = 7.5 (4.3)</td>
<td>-0.40 [ -1.75, 0.95 ]</td>
<td></td>
</tr>
</tbody>
</table>
Analysis 1.6.
Comparison 1 Arnica versus ibuprofen, Outcome 6 Cumulative dose of analgesics (acetaminophen mg) over 3 weeks.
## Analysis 1.7.

Comparison 1 Arnica versus ibuprofen, Outcome 7 Participants (n) reported adverse events.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Arnica gel</th>
<th>Ibuprofen gel</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Widrig 2007</td>
<td>14/105</td>
<td>8/99</td>
<td>1.65 [0.72, 3.76]</td>
</tr>
</tbody>
</table>
Analysis 2.1.
Comparison 2 Capsaicin 0.0125% versus placebo, Outcome 1 Pain VAS 0–100.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Capsicum gel N</th>
<th>Mean(SD)</th>
<th>Placebo N</th>
<th>Mean(SD)</th>
<th>Mean Difference IV(Random,95% CI)</th>
<th>Mean Difference IV(Random,95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kosuwon 2010</td>
<td>65</td>
<td>44.6 (14.6)</td>
<td>34</td>
<td>45.6 (13.5)</td>
<td>-1.00 [-6.76, 4.76]</td>
<td>-1.00 [-6.76, 4.76]</td>
</tr>
</tbody>
</table>
Analysis 2.2.
Comparison 2 Capsaicin 0.0125% versus placebo, Outcome 2WOMAC 0–4 (Overall).
**Analysis 2.3.**
Comparison 2 Capsaicin 0.0125% versus placebo, Outcome 3 Adverse event episodes (n) reported.
**Analysis 3.1.**
Comparison 3 Comfrey versus placebo, Outcome 1 Pain VAS 0–100.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Comfrey</th>
<th>Placebo</th>
<th>Mean (SD)</th>
<th>Mean Difference</th>
<th>IV/Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grube 2007</td>
<td>110</td>
<td>110</td>
<td>42 (28)</td>
<td>83.5 (24)</td>
<td>-41.50 [-48.39, -34.61]</td>
</tr>
</tbody>
</table>

Cameron and Chrubasik

Cochrane Database Syst Rev. Author manuscript; available in PMC 2014 July 21.
Analysis 3.2.
Comparison 3 Comfrey versus placebo, Outcome 2 Pain VAS 0–100 change from baseline.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Comfrey</th>
<th>Placebo</th>
<th>Mean Difference</th>
<th>IV Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Mean(SD)</td>
<td>N</td>
<td>Mean(SD)</td>
<td></td>
</tr>
<tr>
<td>Grube 2007</td>
<td>110</td>
<td>-51.6 (0)</td>
<td>110</td>
<td>-10.1 (0)</td>
</tr>
</tbody>
</table>

Favours comfrey    Favours placebo

Cochrane Database Syst Rev. Author manuscript; available in PMC 2014 July 21.
**Analysis 3.3.**
Comparison 3 Comfrey versus placebo, Outcome 3 Pain VAS 0–100 (at rest) change from baseline.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Comfrey N</th>
<th>Comfrey Mean(SD)</th>
<th>Placebo N</th>
<th>Placebo Mean(SD)</th>
<th>Mean Difference IVRandom,95% CI</th>
<th>Mean Difference IVRandom,95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grube 2007</td>
<td>110</td>
<td>-20.9 (0)</td>
<td>110</td>
<td>-4.6 (0)</td>
<td>0.0 [ 0.0, 0.0 ]</td>
<td></td>
</tr>
</tbody>
</table>
Analysis 3.4.
Comparison 3 Comfrey versus placebo, Outcome 4 Pain VAS 0–100 (movement) change from baseline.
### Analysis 3.5.
Comparison 3 Comfrey versus placebo, Outcome 5 WOMAC-VAS (Pain) change from baseline.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Comfrey</th>
<th>Placebo</th>
<th>Mean Difference</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grube 2007</td>
<td>110</td>
<td>110</td>
<td>12.1 (0)</td>
<td>2.7 (0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0 [0.0, 0.0]</td>
</tr>
</tbody>
</table>

Favours comfrey
Favours placebo
### Analysis 3.6.
Comparison 3 Comfrey versus placebo, Outcome 6 WOMAC-VAS (Stiffness) change from baseline.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Comfrey</th>
<th>Placebo</th>
<th>Mean Difference 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean(SD)</td>
<td>IV/Random, 95% CI</td>
</tr>
<tr>
<td>Grube 2007</td>
<td>110</td>
<td>4.8 (0)</td>
<td>110 1.2 (0) 0.0 [0.0, 0.0]</td>
</tr>
</tbody>
</table>

-100 -50 0 50 100
Favours comfrey Favours placebo
**Analysis 3.7.**
Comparison 3 Comfrey versus placebo, Outcome 7WOMAC-VAS (Function) change from baseline.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Comfrey</th>
<th>Placebo</th>
<th>Mean Difference</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean(SD)</td>
<td>N</td>
<td>Mean(SD)</td>
</tr>
<tr>
<td>Grube 2007</td>
<td>110</td>
<td>43.4 (0)</td>
<td>110</td>
<td>10.7 (0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0 [ 0.0, 0.0 ]</td>
</tr>
</tbody>
</table>

Cameron and Chrubasik

Cochrane Database Syst Rev. Author manuscript; available in PMC 2014 July 21.
**Analysis 3.8.**  
Comparison 3 Comfrey versus placebo, Outcome 8WOMAC-VAS (Overall) change from baseline.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Comfrey</th>
<th>Placebo</th>
<th>Mean Difference</th>
<th>IV/Random,95% CI</th>
<th>IV/Random,95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Mean(SD)</td>
<td>N</td>
<td>Mean(SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grube 2007</td>
<td>110</td>
<td>110</td>
<td>60.4 (0)</td>
<td>14.7 (0)</td>
<td>0.0 [ 0.0, 0.0 ]</td>
</tr>
</tbody>
</table>

Favours comfrey Favours placebo
**Analysis 3.9.**
Comparison 3 Comfrey versus placebo, Outcome 9 Change in SF-36 physical component summary score.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Comfrey</th>
<th>Placebo</th>
<th>Mean Difference</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean(SD)</td>
<td>N</td>
<td>Mean(SD)</td>
</tr>
<tr>
<td>Grube 2007</td>
<td>110</td>
<td>11.9 (0)</td>
<td>110</td>
<td>1.3 (0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0 [0.0, 0.0]</td>
</tr>
</tbody>
</table>
Analysis 3.10.
Comparison 3 Comfrey versus placebo, Outcome 10 Change in SF-36 mental component summary score.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Comfrey</th>
<th>Placebo</th>
<th>Mean Difference</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Mean(SD)</td>
<td>N</td>
<td>Mean(SD)</td>
<td>IV(Random,95% CI)</td>
</tr>
<tr>
<td>Grube 2007</td>
<td>110</td>
<td>110</td>
<td>0.0 [0.0, 0.0]</td>
<td>-10 -5 0 5 10</td>
</tr>
</tbody>
</table>

Review: Topical herbal therapies for treating osteoarthritis
Comparison: 3 Comfrey versus placebo
Outcome: 10 Change in SF-36 mental component summary score
## Analysis 3.11.
Comparison 3 Comfrey versus placebo, Outcome 11 Participants (n) reported adverse events.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Comfrey n/N</th>
<th>Placebo n/N</th>
<th>Risk Difference H,Random,95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grube 2007</td>
<td>7/110</td>
<td>15/110</td>
<td>-0.07 [-0.15, 0.01]</td>
</tr>
</tbody>
</table>

Cameron and Chrubasik. Analysis 3.11. Comparison 3 Comfrey versus placebo, Outcome 11 Participants (n) reported adverse events.
Analysis 4.1.
Comparison 4 Marhame-Mafasel versus placebo, Outcome 1 WOMAC-VAS (Pain) change from baseline.
### Analysis 4.2.
Comparison 4 Marhame-Mafasel versus placebo, Outcome 2WOMAC-VAS (Stiffness) change from baseline.
Analysis 4.3.
Comparison 4 Marhame-Mafasel versus placebo, Outcome 3WOMAC-VAS (Function) change from baseline.
Analysis 4.4.
Comparison 4 Marhame-Mafasel versus placebo, Outcome 4WOMAC-VAS (Overall) change from baseline.
Analysis 4.5.
Comparison 4 Marhame-Mafasel versus placebo, Outcome 5 Participants (n) reporting adverse events.
Analysis 5.1.
Comparison 5 Stinging nettle versus placebo, Outcome 1 WOMAC 0–4 (Pain) at 1 week.
Analysis 5.2.
Comparison 5 Stinging nettle versus placebo, Outcome 2 WOMAC 0–4 (Stiffness) at 4 weeks.
Analysis 5.3.
Comparison 5 Stinging nettle versus placebo, Outcome 3 WOMAC 0–4 (Function) at 4 weeks.
Analysis 5.4.
Comparison 5 Stinging nettle versus placebo, Outcome 4 Participants (n) reported adverse events.
### Analysis 6.1.
Comparison 6 FNZG versus placebo, Outcome 1 Pain on walking VAS 0–100.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>FNZG patch</th>
<th>Placebo</th>
<th>Mean Difference IV/Fixed</th>
<th>95% CI</th>
<th>Weight</th>
<th>Mean Difference IV/Fixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang 2012</td>
<td>34.19 (17.85)</td>
<td>35.63 (17.92)</td>
<td>-1.44</td>
<td>[-9.28, 6.40]</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td><strong>60</strong></td>
<td><strong>30</strong></td>
<td><strong>-1.44</strong></td>
<td><strong>[-9.28, 6.40]</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: not applicable
Test for overall effect: Z = 0.36 (P = 0.72)
Test for subgroup differences: Not applicable
Analysis 6.2.
Comparison 6 FNZG versus placebo, Outcome 2WOMAC 0–4 (Pain).

Cochrane Database Syst Rev. Author manuscript; available in PMC 2014 July 21.
Analysis 6.3.
Comparison 6 FNZG versus placebo, Outcome 3WOMAC 0–4 (Stiffness).
Analysis 6.4.
Comparison 6 FNZG versus placebo, Outcome 4WOMAC 0–4 (Function).
Analysis 6.5.
Comparison 6 FNZG versus placebo, Outcome 5 WOMAC 0–4 (Overall).

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>FNZG patch</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean(SD)</td>
</tr>
<tr>
<td>Wang 2012</td>
<td>60</td>
<td>29.71 (20.09)</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>60</td>
<td>30</td>
</tr>
</tbody>
</table>

Heterogeneity: not applicable
Test for overall effect: Z = 0.87 (P = 0.38)
Test for subgroup differences: Not applicable
Analysis 6.6.
Comparison 6 FNZG versus placebo, Outcome 6 Participants (n) reported adverse events.
Analysis 7.1.
Comparison 7 SJG versus placebo, Outcome 1 Pain on walking VAS 0–100.
### Analysis 7.2.
Comparison 7 SJG versus placebo, Outcome 2WOMAC 0–4 (Pain).

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>SJG N</th>
<th>Mean(SD)</th>
<th>Placebo N</th>
<th>Mean(SD)</th>
<th>Mean Difference</th>
<th>Weight</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang 2012</td>
<td>60</td>
<td>6.79 (3.41)</td>
<td>30</td>
<td>8.59 (4.47)</td>
<td>0.00</td>
<td>100.0%</td>
<td>-1.80 [-3.62, 0.02]</td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td><strong>60</strong></td>
<td><strong>30</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>100.0%</strong></td>
<td><strong>-1.80 [-3.62, 0.02]</strong></td>
</tr>
</tbody>
</table>

Heterogeneity: not applicable
Test for overall effect: Z = 1.94 (P = 0.052)
Test for subgroup differences: Not applicable
### Analysis 7.3.
Comparison 7 SJG versus placebo, Outcome 3WOMAC 0–4 (Stiffness).

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>SJG</th>
<th>Placebo</th>
<th>Mean Difference</th>
<th>Weight</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang 2012</td>
<td>60</td>
<td>30</td>
<td></td>
<td>100.0%</td>
<td>-0.37 [-1.19, 0.45]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>60</td>
<td>30</td>
<td></td>
<td>100.0%</td>
<td>-0.37 [-1.19, 0.45]</td>
</tr>
</tbody>
</table>

Heterogeneity: not applicable
Test for overall effect: Z = 0.88 (P = 0.38)
Test for subgroup differences: Not applicable
Analysis 7.4.
Comparison 7 SJG versus placebo, Outcome 4WOMAC 0–4 (Function).
Analysis 7.5.
Comparison 7 SJG versus placebo, Outcome 5WOMAC 0–4 (Overall).
Analysis 7.6.
Comparison 7 SJG versus placebo, Outcome 6 Participants (n) reported adverse events.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>SJG  n/N</th>
<th>Placebo n/N</th>
<th>Weight</th>
<th>Odds Ratio M-H,Fixed 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang 2012</td>
<td>4/60</td>
<td>0/30</td>
<td>100.0%</td>
<td>4.86 [0.25, 93.27]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>60</td>
<td>30</td>
<td>100.0%</td>
<td>4.86 [0.25, 93.27]</td>
</tr>
</tbody>
</table>

Total events: 4 (SJG), 0 (Placebo)
Heterogeneity: not applicable
Test for overall effect: Z = 1.05 (P = 0.29)
Test for subgroup differences: Not applicable
Analysis 8.1.
Comparison 8 FNZG versus SJG, Outcome 1 Pain on walking VAS 0–100.
### Analysis 8.2.

Comparison 8 FNZG versus SJG, Outcome 2WOMAC 0–4 (Pain).

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Experimental</th>
<th>Control</th>
<th>Mean Difference</th>
<th>Weight</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang 2012</td>
<td>60</td>
<td>60</td>
<td></td>
<td></td>
<td>-0.66 [ -0.73, 2.05 ]</td>
</tr>
</tbody>
</table>

Total (95% CI) 60 60 100.0% 0.66 [ -0.73, 2.05 ]
Analysis 8.3.
Comparison 8 FNZG versus SJG, Outcome 3 WOMAC 0–4 (Stiffness).
Analysis 8.4.
Comparison 8 FNZG versus SJG, Outcome 4WOMAC 0–4 (Function).
### Analysis 8.5.
Comparison 8 FNZG versus SJG, Outcome 5WOMAC 0–4 (Overall).

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>FNZG</th>
<th>SJG</th>
<th>Mean difference</th>
<th>Weight</th>
<th>Mean difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang 2012</td>
<td>60</td>
<td>60</td>
<td>-5.74 (6.8)</td>
<td>100.0%</td>
<td>0.89 [-5.74, 7.52]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>60</td>
<td>60</td>
<td></td>
<td>100.0%</td>
<td>0.89 [-5.74, 7.52]</td>
</tr>
</tbody>
</table>

Heterogeneity: not applicable
Test for overall effect: Z = 0.26 (P = 0.79)
Test for subgroup differences: Not applicable
### Analysis 8.6.

Comparison 8 FNZG versus SJG, Outcome 6 Participants (n) reported adverse events.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>FNZG n/N</th>
<th>SJG n/N</th>
<th>Odds Ratio M-H, Fixed 95% CI</th>
<th>Weight</th>
<th>Odds Ratio M-H, Fixed 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang 2012</td>
<td>5/60</td>
<td>4/60</td>
<td></td>
<td>100.0%</td>
<td>1.27, 0.32, 4.99</td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td><strong>60</strong></td>
<td><strong>60</strong></td>
<td></td>
<td><strong>100.0%</strong></td>
<td><strong>1.27, 0.32, 4.99</strong></td>
</tr>
</tbody>
</table>

Test for overall effect: Z = 0.35 (P = 0.73)
Test for subgroup differences: Not applicable
Table 1

Commission E approved monographs of medicinal plants for musculoskeletal complaints

<table>
<thead>
<tr>
<th>Species</th>
<th>Plant part</th>
<th>Preparations</th>
<th>Dosage</th>
<th>Treatment duration</th>
<th>Adverse events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arnica montana</td>
<td>flower (fresh or dried)</td>
<td>tincture ointment oil</td>
<td>3–10 times diluted 20–25% tincture or 15% oil 1:4 fatty oil</td>
<td>skin reactions</td>
<td></td>
</tr>
<tr>
<td>Arnica chamissonis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capsicum species¹</td>
<td>dried fruit</td>
<td>extracts with capsai-cinoids</td>
<td>semi-liquid: 0.02–0.05% capsaicinoids</td>
<td>2 days</td>
<td>skin reactions; local hyperaemic and nerve damaging</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>liquid: 0.005–0.01% capsaicinoids</td>
<td>2–14 days</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>poultices: 10–40µg capsacinoids/cm²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mentha arvensis</td>
<td>fresh flowering herb</td>
<td>essential oil²</td>
<td>several drops rubbed into the skin 5–20% oil or semisolid preparations</td>
<td>2–14 days</td>
<td></td>
</tr>
<tr>
<td>Mentha piperita</td>
<td>fresh flowering sprigs</td>
<td>essential oil³</td>
<td>several drops rubbed into the skin dilutions in oil or semisolid preparations</td>
<td></td>
<td>skin reactions</td>
</tr>
<tr>
<td>Pica species⁴</td>
<td>tips of branches</td>
<td>essential oil</td>
<td>not stated</td>
<td>skin reactions</td>
<td></td>
</tr>
<tr>
<td>Pinus species⁵</td>
<td>fresh boughs with needles and tips</td>
<td>5–50% essential oil</td>
<td>several drops rubbed into the skin dilutions in oil or semisolid preparations</td>
<td></td>
<td>skin reactions</td>
</tr>
<tr>
<td>Symphytum officinale</td>
<td>herb and leaf</td>
<td>ointment</td>
<td>5–20% dried drug⁶</td>
<td>restricted to 4–6 weeks</td>
<td>None known</td>
</tr>
<tr>
<td></td>
<td>root</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urtica dioica</td>
<td>herb and leaf (fresh)</td>
<td>crude material</td>
<td>5–20% fresh or dried drug⁶</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urtica urens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ eg: C. fructescens.
² 3–17% menthyl acetate, 42% menthol, 40% menthone.
³ 4.5–10% menthyl acetate, 44% menthol, 15–32% menthone.
⁴ P. arbies, P. excelsa, P. alba, P. sachalinensis, P. sibirica.
⁵ P. sylvestris, P. mugo, P. nigra.
⁶ with less that 100 µg pyrrolizidin alkaloids/day.
## Table 2

ESCOP monographs for musculoskeletal complaints based on experimental and clinical studies

<table>
<thead>
<tr>
<th>Species</th>
<th>Starting material</th>
<th>Preparations</th>
<th>Dosage</th>
<th>Treatment duration</th>
<th>Adverse events</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arnica montana</strong></td>
<td>flower with 4% helenalin (sesquiterpenes lactones)</td>
<td>tincture ointment cream gel compress</td>
<td>dilutions from 5–25% tincture or fluid extract: 1/3 to 1:10</td>
<td></td>
<td>skin reactions</td>
</tr>
<tr>
<td><strong>Capsicum species</strong></td>
<td>fruit (dried) with not less than 0.4% capsaicinoids</td>
<td>liquid</td>
<td>0.025-0.075% capsaicinoids</td>
<td>3 weeks</td>
<td>skin reactions local hyperaemic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>semiliquid</td>
<td>poultice: 10–40 μg capsaicinoids cm⁻²</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mentha piperita</strong></td>
<td>fresh flowering herb</td>
<td>essential oil 1.25–16%²</td>
<td>several drops rubbed into the skin dilutions in semiliquid preparations</td>
<td></td>
<td>skin reactions</td>
</tr>
<tr>
<td><strong>Symphytum officinale</strong></td>
<td>root with up to 4.7% allantoin</td>
<td>ointment</td>
<td>35% root extract (DER 1.2, solvent ethanol 60%) 3–4/day</td>
<td>restricted in some EU countries</td>
<td>None reported</td>
</tr>
<tr>
<td><strong>Urtica dioica</strong></td>
<td>fresh herb or leaf</td>
<td>crude material</td>
<td>30 seconds / daily</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Urtica urens</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ eg: *C. frutescens, C. annuum.*

² 30–55% menthol, 14–32% menthone, 2.8–10% methyl acetate, 1–9% menthofuran, 3.5–14% cineol, etc.
Table 3

Herbal medicinal products used for the treatment of osteoarthritis

<table>
<thead>
<tr>
<th>Botanical name</th>
<th>Plant part/s</th>
<th>Tradename</th>
<th>Preparation</th>
<th>Drug/Extract</th>
<th>mg/day</th>
<th>Constituent marker</th>
<th>Quantity of marker</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arnica montana (local)</td>
<td>herb</td>
<td>A. Vogel Arnica Gel</td>
<td>tincture, 50% ethanol(^1)</td>
<td>20:1</td>
<td>3 × 4 cm</td>
<td>not stated</td>
<td>not stated</td>
<td>Widrig 2007.</td>
</tr>
<tr>
<td>Capsicum species</td>
<td>fruit</td>
<td>Capsika Gel</td>
<td>extract, ethanol (concentration not stated)</td>
<td>0.0125% TID × 2 inches gel</td>
<td>capsaicin</td>
<td>0.365</td>
<td>Kosuwon 2010</td>
<td></td>
</tr>
<tr>
<td>Symphytum officinale (local)</td>
<td>root</td>
<td>Kytt Salbe f</td>
<td>ethanolic (60%) extract</td>
<td>2:1</td>
<td>6 (3 × 2mg)</td>
<td>allantoin</td>
<td>0.2–0.5%</td>
<td>Grube 2007.</td>
</tr>
<tr>
<td>Arnabia euchroma + Matricaria chamomilla + other medicinal plant parts</td>
<td>not stated</td>
<td>Marhame-Mafasel</td>
<td>not stated</td>
<td>not stated</td>
<td>4500mg</td>
<td>not stated</td>
<td>not stated</td>
<td>Soltanian 2010</td>
</tr>
<tr>
<td>rheum Arisematis, radix Aconiti, flos Caryophylli, cortex Cinnamomi, radix Angelicae dehariae, herba Asari, rhizoma Chuanxiong, radix Cynanchi paniculati, Olbanum, Myrrha, Camphora, Bornedum synthetcum.</td>
<td>see previous column</td>
<td>Fuang Nanxing Zhitong Gao (FNZG)(2)</td>
<td>not stated</td>
<td>not stated</td>
<td>not stated</td>
<td>hypa-conitine (C33H45NO10); eugenol (C10H12O2)</td>
<td>not stated</td>
<td>Wang 2012</td>
</tr>
<tr>
<td>rheum Arisematis, radix Aconiti, radix Angelicae Dahariae, cortex Cinnamomi, Camphora, Borneol Synthetcum, radix Angelicae Puhensensis, cortex Acanthopanicis, rhizoma Curcuma Longae, flos Carthami, folium Artemisiae argyi, rhizoma Atractylodis, rhizoma Pinellia, semen Sinapis, semen Vaccariae, radix Aconiti kusnezoffii, herba Menthae.</td>
<td>see previous column</td>
<td>Shangshi Jietong Gao (SJG)(2)</td>
<td>not stated</td>
<td>not stated</td>
<td>not stated</td>
<td>hypa-conitine (C33H45NO10); eugenol (C10H12O2)</td>
<td>not stated</td>
<td>Wang 2012</td>
</tr>
</tbody>
</table>

\(^1\) Information provided by manufacturer but not reported in paper.

\(^2\) Jiangsu Nanxing Pharmaceutical Company.
Table 4

Summary of findings: FNZG patches for osteoarthritis of the knee

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain</strong></td>
<td>Control</td>
<td>Pain on walking VAS 0–100</td>
<td>The mean pain in the intervention groups was 1.44 lower (9.28 lower to 6.4 higher)</td>
<td>90 (1 study)</td>
<td>💬💼 moderate(^1),(^2).</td>
</tr>
<tr>
<td></td>
<td>FNZG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Function</strong></td>
<td>Control</td>
<td>WOMAC 0–4 (Function)</td>
<td>The mean function in the intervention groups was 2.61 lower (9.5 lower to 4.28 higher)</td>
<td>90 (1 study)</td>
<td>💬💼 moderate(^1),(^2).</td>
</tr>
<tr>
<td></td>
<td>FNZG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adverse events</strong></td>
<td>Study population</td>
<td>Participants (n) reported adverse events.</td>
<td>OR 6.05 (0.32 to 113.05)</td>
<td>90 (1 study)</td>
<td>💬💼 moderate(^1),(^2).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Follow-up: mean 7 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0/30</td>
<td>Adverse events</td>
<td>Participants (n) withdrew due to adverse effects</td>
<td>Not estimable</td>
<td>90 (1 study)</td>
</tr>
<tr>
<td></td>
<td>1/60</td>
<td></td>
<td>Follow-up: mean 7 days</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adverse events</td>
<td>Participants (n) reported serious adverse events</td>
<td>Not estimable</td>
<td>90 (1 study)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Follow-up: mean 7 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Quality of Life</strong></td>
<td>Not measured or reported.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The basis for the assumed risk (e.g., the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio; OR: Odds ratio.

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

\(^1\) Single study.
Table 5

Summary of findings: SJG patches for osteoarthritis of the knee

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assumed risk</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Corresponding risk</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SJG</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pain</strong></td>
<td>Pain on walking VAS 0–100 Follow-up: mean 7 days</td>
<td>The mean pain in the intervention groups was 1.08 higher (6.24 lower to 8.4 higher)</td>
<td>90 (1 study)</td>
<td>mod. 1,2</td>
<td></td>
</tr>
<tr>
<td><strong>Function</strong></td>
<td>Function WOMAC 0–4 (Function) Follow-up: mean 7 days</td>
<td>The mean function in the intervention groups was 2.97 lower (9.6 lower to 3.66 higher)</td>
<td>90 (1 study)</td>
<td>mod. 1,2</td>
<td></td>
</tr>
<tr>
<td><strong>Adverse events</strong></td>
<td>Study population</td>
<td>OR 4.86 (0.25 to 93.27)</td>
<td>90 (1 study)</td>
<td>mod. 1,2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adverse events Participants (n) reported adverse events Follow-up: mean 7 days</td>
<td>0 per 1000 0 per 1000 (0 to 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adverse events Participants (n) withdrew due to adverse effects 7 days</td>
<td>0/30 0/60</td>
<td>Not estimable</td>
<td>90 (1 study)</td>
<td>mod. 1,2</td>
</tr>
<tr>
<td></td>
<td>Adverse events Participants (n) reported serious adverse events Follow-up: mean 7 days</td>
<td>0/30 0/60</td>
<td>Not estimable</td>
<td>90 (1 study)</td>
<td>mod. 1,2</td>
</tr>
<tr>
<td><strong>Quality of Life</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio; OR: Odds ratio.

GRADE Working Group grades of evidence

**High quality:** Further research is very unlikely to change our confidence in the estimate of effect.

**Moderate quality:** Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

**Low quality:** Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

**Very low quality:** We are very uncertain about the estimate.

Single study.
### Table 6

Summary of findings: FNZG patches versus SJG patches for osteoarthritis of the knee

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient or population:</strong> patients with osteoarthritis of the knee</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Settings:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intervention:</strong> FNZG versus SJG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain on walking VAS 0–100</td>
<td>SJG: The mean pain in the intervention groups was 2.52 lower (8.24 lower to 3.2 higher)</td>
<td></td>
<td>120 (1 study)</td>
<td>★★★☆ moderate</td>
<td></td>
</tr>
<tr>
<td>Follow-up: mean 7 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAC 0–4 (Function)</td>
<td>SJG: The mean function in the intervention groups was 0.36 higher (4.49 lower to 5.21 higher)</td>
<td></td>
<td>120 (1 study)</td>
<td>★★★☆ moderate</td>
<td></td>
</tr>
<tr>
<td>Follow-up: mean 7 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse events</td>
<td>Study population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants (n) reported adverse events</td>
<td>67 per 1000</td>
<td>83 per 1000 (22 to 263)</td>
<td>OR 1.27 (0.32 to 4.99)</td>
<td>120 (1 study)</td>
<td>★★★☆ moderate</td>
</tr>
<tr>
<td>Follow-up: mean 7 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse events</td>
<td>Study population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants (n) withdrew due to adverse effects</td>
<td>17 per 1000</td>
<td>0 per 1000 (0 to 0)</td>
<td>Not estimable</td>
<td>120 (1 study)</td>
<td>★★★☆ moderate</td>
</tr>
<tr>
<td>Follow-up: mean 7 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse events</td>
<td>Study population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants (n) reported serious adverse events</td>
<td>0/60</td>
<td>0/60</td>
<td>Not estimable</td>
<td>120 (1 study)</td>
<td>★★★☆ moderate</td>
</tr>
<tr>
<td>Follow-up: mean 7 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Quality of Life</strong></td>
<td>Not measured or reported.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio; OR: Odds ratio.

GRADE Working Group grades of evidence

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

1 Single study.
Confirmatory design, statistical power 80%, alpha 0.05.
Table 7
Summary of findings: Marhame-Mafasel for osteoarthritis of the knee

<table>
<thead>
<tr>
<th>Marhame-Mafasel for osteoarthritis of the knee</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain due to OA: change from baseline</td>
<td>Control: The mean pain due to OA: change from baseline in the intervention groups was 5.62 lower (17.84 lower to 6.6 higher)</td>
<td>42 (1 study)</td>
<td>☒ ☒ ☒ low²,³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Function due to OA: change from baseline</td>
<td>Function: The mean function due to OA: change from baseline in the intervention groups was 1.09 lower (9.4 lower to 7.22 higher)</td>
<td>42 (1 study)</td>
<td>☒ ☒ ☒ low²,³</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adverse events</strong></td>
<td>Study population</td>
<td>RR 5.00 (0.25 to 98.27)</td>
<td>42 (1 study)</td>
<td>☒ ☒ ☒ low²,³</td>
<td></td>
</tr>
<tr>
<td>Participants (n) withdrew due to adverse effects</td>
<td>0 per 1000 (0 to 0)</td>
<td>OR 5.51 (0.25 to 122.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants (n) reported serious adverse events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of Life</td>
<td>Not measured or reported.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

GRADE Working Group grades of evidence

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

¹ Crossover trial: 3 week interventions arms, without washout period.
² Single study.
3 Exploratory study design; power, effect, and sample size not determined *a priori.*
Table 8
Summary of findings: stinging nettle for osteoarthritis of the thumb

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks*</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assumed risk</td>
<td>Corresponding risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Stinging nettle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td>The mean pain in the control groups was 37.64 mm</td>
<td>The mean pain in the intervention groups was 13.37 lower (0 to 0 higher)</td>
<td>54 (1 study)</td>
<td>+++++ very low¹ ² ³</td>
<td>Crossover trial: Participants (n) with adverse events reported for whole trial only (ie: both arms of crossover combined)</td>
</tr>
<tr>
<td>Function</td>
<td>Not measured or reported.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse events</td>
<td>Participants (n) reported adverse events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse events</td>
<td>Participants (n) withdrew due to adverse effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse events</td>
<td>Participants (n) reported serious adverse events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of life</td>
<td>Health Assessment Questionnaire (HAQ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval.

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

¹ Single study.
2 Criteria for diagnosis of OA not specified at baseline. Other arthritides are possible confounders.

3 Most outcome data were reported as means only (not standard deviations) without confidence intervals. Data for first and second periods were aggregated and are insufficient for extraction from the first arm (up to crossover) only. We have presented aggregated data for pain (VAS 0–100) after 1 week of intervention and re-calculated the mean difference between groups at this time point. We note that the mean difference between groups is –13.37, not 15.08, which is the sum of the within-groups mean changes reported by the authors as mean difference.
Table 9

Summary of findings: stinging nettle for osteoarthritis of the knee

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain</strong></td>
<td>Assumed risk Stinging nettle</td>
<td>RR 2.00 (0.20 to 20.41)</td>
<td>48 (1 study)</td>
<td>😓很低/2,3,4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control Stinging nettle</td>
<td>OR 2.11 (0.18 to 25.17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WOMAC 0–4 (Pain; higher means worse)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Follow-up: mean 1 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The mean pain due to OA in the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>intervention groups was 2 higher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.19 to 3.81 higher)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Function</strong></td>
<td>Assumed risk Stinging nettle</td>
<td></td>
<td></td>
<td>😓很低/2,3,4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control Stinging nettle</td>
<td>RR 4.00 (0.12 to 137)</td>
<td>39 (1 study)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WOMAC 0–4 (Function; higher means worse)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Follow-up: mean 4 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The mean function due to OA in the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>intervention groups was 5 higher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.9 to 9.1 higher)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adverse events</strong></td>
<td>Participants (n) reported adverse</td>
<td>RR 2.00 (0.20 to 20.41)</td>
<td>48 (1 study)</td>
<td>😓很低/2,3,4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>events</td>
<td>OR 2.11 (0.18 to 25.17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Follow-up: mean 1 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participants (n) withdrew due to</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>adverse effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adverse events</strong></td>
<td>Participants (n) reported serious</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>adverse events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Follow-up: mean 1 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participants (n) withdrew due to</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>adverse effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Quality of life</strong></td>
<td>Not measured or reported.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; OR: Odds ratio.

GRADE Working Group grades of evidence
- High quality: Further research is very unlikely to change our confidence in the estimate of effect.
- Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect.
- Low quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
- Very low quality: We are very uncertain about the estimate.
1 Diagnosis not consistent with ACR criteria. OA not distinguished from other causes of knee pain in older adults at baseline (presumptive diagnosis). Although OA is the most likely cause of knee pain in older adults, other arthritides are possible confounders.

2 Exploratory study design; power, effect, and sample size not determined a priori.

3 Placebo validity and blinding may be compromised by stinging effect of active intervention.

4 Patients were follow up for pain and function measures at 4 and 16 weeks. In both intervention and control groups, further improvements in these domains were identified at follow up time points. These long term improvements suggest that short term changes in pain and function represent normal variation within individuals with OA, and are not of clinical importance.

5 1 week intervention, 15 weeks follow up.

6 Urtica dioica stimulates pain receptors in the skin, which is part of the effect of this topical agent. “Stinging sensations” were not included in reported adverse events, probably because participants understood that this effect was likely with the intervention. One participant in the control group reported stiffness in the knee as an adverse event. Stiffness is an expected outcome of untreated OA. These results suggest that insufficient blinding to the intervention may have confounded the reporting of adverse events.
### SUMMARY OF FINDINGS FOR THE MAIN COMPARISON

**Arnica versus ibuprofen for osteoarthritis of the hand**

**Patient or population:** patients with osteoarthritis of the hand  
**Settings:** Community, Switzerland  
**Intervention:** Arnica montana  
**Comparison:** Ibuprofen

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain</strong></td>
<td></td>
<td></td>
<td><strong>Ibuprofen</strong></td>
<td><strong>Arnica</strong></td>
<td></td>
</tr>
<tr>
<td>VAS 0 to 100 (higher scores means worse)</td>
<td>The mean pain in the control group was 44.2 points on a 100 point scale</td>
<td>The mean pain in the intervention group was <strong>3.8 lower</strong> (10.1 lower to 2.5 higher)</td>
<td>174 (1 study)</td>
<td>💡 moderate 1,2,3</td>
<td>Absolute reduction in pain was 4% (10% reduction to 3% increase); relative reduction in pain 5% (15% reduction to 4% increase); NNT n/a.4</td>
</tr>
<tr>
<td>Function</td>
<td>Hand algofunctional index (higher scores means worse)</td>
<td>The mean function in the control group was 7.5 points on a 30 point scale</td>
<td>The mean hand function in the intervention group was <strong>0.4 lower</strong> (1.75 lower to 0.95 higher)</td>
<td>174 (1 study)</td>
<td>💡 moderate 1,2,3</td>
</tr>
</tbody>
</table>

| Adverse events | Study population | RR 1.65 (0.72 to 3.76) | 204 (1 study) | 💡 moderate 1,2,3 | Absolute risk of adverse events was 5% higher in the Arnica group (3% lower to 14% higher); NNT n/a.4 |

<table>
<thead>
<tr>
<th>Adverse events</th>
<th>Participants (n) reported events</th>
<th>Not estimable</th>
<th>Reported NIL withdrawal due to adverse events.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adverse events</td>
<td>Participants (n) reported serious adverse events</td>
<td>Not estimable</td>
<td>Reported NIL serious adverse events.5</td>
</tr>
</tbody>
</table>

**Quality of life**  
Quality of Life not measured.

*The basis for the **assumed risk** (e.g., the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; OR: Odds ratio.

**GRADE Working Group grades of evidence**

**High quality:** Further research is very unlikely to change our confidence in the estimate of effect.

**Moderate quality:** Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

**Low quality:** Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

**Very low quality:** We are very uncertain about the estimate.

1 Single study. Treatment effect crosses midline (no effect).
Non-inferiority hypothesis: that *Arnica* is not inferior to ibuprofen for the treatment of osteoarthritis of the hand.

Confirmatory design, statistical power 80%, alpha 0.024.

Number needed to treat (NNT) = not applicable (n/a) when result is not statistically significant. NNT for continuous outcomes calculated using Wells Calculator (CMSG editorial office). NNT for dichotomous outcomes calculated using Cates NNT calculator ([http://www.mrontline.net/visualrx/](http://www.mrontline.net/visualrx/)).

Negligible percentage change less than 1%.

Reported one case of back pain due to a fall, leading to withdrawal from the study; this event is neither withdrawal due to adverse event, nor a serious adverse event, as defined for this review.
### Capsicum for osteoarthritis of the knee

**Patient or population:** patients with osteoarthritis of the knee  
**Settings:** Community, Thailand  
**Intervention:** Capsicum extract

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
</table>
| **Pain** | VAS 0–100 (higher means worse)  
Follow-up: mean 4 weeks | The mean pain in the control group was 45.6 points on a 100 point scale  
The mean pain in the intervention group was 1 lower (6.7 lower to 4.7 higher) | 99 (1 study) | ★★★★ moderate | Absolute reduction in pain was 1% (7% reduction to 5% increase); relative reduction in pain 2% (10% reduction to 7% increase); NNT n/a. |
| **Function** | WOMAC 0–4 (Function; higher means worse)  
Follow-up: mean 4 weeks | Measured, but not reported.  
The mean function in the control group was 34.79 on a 96 point scale  
The mean function in the intervention group was 2.64 lower (9.51 lower to 4.23 higher) | 99 (1 study) | ★★★★ moderate | Absolute functional improvement 3% (10% improvement to 4% decline); relative functional improvement 5% (19% improvement to 9% decline); NNT n/a. |
| **Adverse events** | Event episodes (n) reported  
Follow-up: mean 4 weeks | Not estimable | 676 (1 study) | ★★★★ moderate | Absolute risk of adverse events was 61% higher in the capsaicin group (55% to 67% higher); NNT = 1.64 (95% CI 1.82 to 1.50) |

**Adverse events**  
Participants (n) withdrew due to adverse effects

**Adverse events**  
Participants (n) reported serious adverse events

**Quality of Life**  
Not measured or reported.

*The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: confidence interval; OR: Odds ratio.

GRADE Working Group grades of evidence  
**High quality:** Further research is very unlikely to change our confidence in the estimate of effect.
### Comfrey for osteoarthritis of the knee

**Patient or population:** patients with osteoarthritis of the knee  
**Settings:** Community, Germany  
**Intervention:** Comfrey

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAS 0–100 (higher means worse) Follow-up: mean 3 weeks</td>
<td>The mean pain in the control group was 83.5 points on 100 point scale.</td>
<td>The mean pain: change from baseline in the intervention groups was 41.5 lower (34 to 48 lower)</td>
<td>220 (1 study)</td>
<td>생활</td>
<td>Absolute reduction in pain was 42% (34% to 48% reduction); relative reduction in pain 48% (36% to 51% reduction); NNT = 1.84 (95% CI 1.7 to 2.1).</td>
</tr>
<tr>
<td>Function: change from baseline</td>
<td>WOMAC VAS (Function; higher means worse) Follow-up: mean 3 weeks</td>
<td>Not estimable</td>
<td>Not estimable</td>
<td>220 (1 study)</td>
<td>생활</td>
</tr>
<tr>
<td>Adverse events</td>
<td>Study population</td>
<td>RR 0.47 (0.20 to 1.10)</td>
<td>220 (1 study)</td>
<td>생활</td>
<td>Absolute risk of adverse events was 7% lower in the comfrey group (15% lower to 1% higher); NNT n/a.</td>
</tr>
<tr>
<td>Adverse events</td>
<td>Participants (n) withdrew due to adverse effects</td>
<td>Not estimable</td>
<td></td>
<td>Withdrawal due to adverse events not reported.</td>
<td></td>
</tr>
<tr>
<td>Adverse events</td>
<td>Participants (n) reported serious adverse events</td>
<td>Not estimable</td>
<td></td>
<td>Reported NIL serious adverse events.</td>
<td></td>
</tr>
<tr>
<td>Quality of life SF-36d</td>
<td></td>
<td></td>
<td>220 (1 study)</td>
<td>생활</td>
<td></td>
</tr>
</tbody>
</table>

*The basis for the assumed risk (e.g., the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; OR: Odds ratio.

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

Single study. Treatment effect crosses midline (no effect).

---

**Moderate quality:** Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.  
**Low quality:** Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.  
**Very low quality:** We are very uncertain about the estimate.
Randomisation and allocation concealment inadequately reported. Per protocol analysis only.

Crossover trial: 4 week intervention arms plus 1 week washout.

Capsicum extract gel stimulates heat receptors in the skin, which is part of the effect of this topical agent. Including “burning sensations” among the reported adverse events may have inflated this outcome. Insufficient blinding to the intervention may have confounded the reporting of adverse events.

Number needed to treat (NNT) = not applicable (n/a) when result is not statistically significant. NNT for continuous outcomes calculated using Wells Calculator (CMSG editorial office). NNT for dichotomous outcomes calculated using Cates NNT calculator (http://www.nnotonline.net/visualrx/).

Standard deviations for pain measures were estimated from graphical data.

Most outcome data reported as change scores, percentages, and graphs only, insufficient for extraction.

Criteria for diagnosis of OA not specified at baseline. Diagnosis not consistent with ACR criteria, but likely to be OA.

SF-36 has a recall period of 4 weeks. Participants are asked to recall their health perceptions “over the last 4 weeks”. Use of this outcome measure to investigate interventions less than 4 weeks duration is likely to be imprecise.
CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

<table>
<thead>
<tr>
<th>Study</th>
<th>Methods</th>
<th>Participants</th>
<th>Interventions</th>
<th>Outcomes</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grube 2007</td>
<td>Randomised, double-blind, placebo control, 2 parallel groups, 2 centre study. Duration 3 weeks</td>
<td>Randomised n=220, Completed n=186. Mean age 58 years. M:F 67:153. Inclusion: OA knee (criteria not specified), pain VAS 0–100 &gt;40mm</td>
<td>Kytta-Salbe® E. Symphytum officinale radix (comfrey root) extract, 6g (2g TID), ointment. Placebo control: ingredients not reported, ointment. Regimen: If bilateral OA, treat both knees, but outcome measures limited to most painful joint only. Massage 6cm long thread of ointment into skin covering the knee three times daily</td>
<td>Pain at rest VAS 0–100, pain with movement VAS 0–100, WOMAC-VAS 0–100 (24 items, 3 subscales; all VAS 0 indicates no deficits, 100 indicates worst possible deficits)</td>
<td>Results favour intervention.</td>
</tr>
</tbody>
</table>

**Risk of bias**

<table>
<thead>
<tr>
<th>Bias</th>
<th>Authors’ judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Low risk</td>
<td>Described as randomised, method not reported(^1). Baseline parameters compared for significant differences.</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Low risk</td>
<td>Allocation concealment not reported(^1).</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias)</td>
<td>Low risk</td>
<td>Described as double-blind. Active intervention and placebo not distinguished by look, taste, smell or packaging</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias)</td>
<td>Low risk</td>
<td>Reported withdrawals. Included full analyses (intention-to-treat) and valid case analyses (per-protocol)</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Unclear risk</td>
<td>Confirmatory design, large sample size, statistical power not reported, alpha 0.05. (low risk) Most outcome data reported as change scores, percentages, and graphs only, insufficient for extraction. (unclear risk) Standard deviations for pain estimated from graphical data. Reported adverse events. (low risk)</td>
</tr>
<tr>
<td>Other bias</td>
<td>Unclear risk</td>
<td>Criteria for diagnosis of OA not specified at baseline. Diagnosis not consistent with ACR criteria, but likely to be OA. (unclear risk) Reported compliance with Declaration of Helsinki and ICH GCP guidelines. (low risk)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study</th>
<th>Methods</th>
<th>Participants</th>
<th>Interventions</th>
<th>Outcomes</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kosuwon 2010</td>
<td>Randomised, double-blind, placebo control, 2 group, crossover study. Duration 4 weeks each arm, 1 week washout</td>
<td>Randomised n=100, Completed n=99. Mean age 61 years, range 44–82 years. M:F = 100:0. Inclusion: OA knee (ACR criteria)</td>
<td>Capsika gel: Capsicum (species not stated) extract. Placebo control: ingredients not reported. Regimen: TID, applied 2 inches of extruded gel around the knee and rubbed in until dry Rescure medication permitted: paracetamol (acetaminophen) up to 1500mg (3 × 500mg)</td>
<td>Pain VAS 0–100 (0 indicates no pain, 100 indicates worst possible pain), WOMAC 0–4 (24 items, 3 subscales, higher scores indicate worse deficits)</td>
<td>Results favour intervention.</td>
</tr>
</tbody>
</table>
Random sequence generation (selection bias) | Unclear risk | Described as randomised, method not reported.
---|---|---
Allocation concealment (selection bias) | Unclear risk | Allocation concealment not reported.
Blinding (performance bias and detection bias) | Low risk | Described as double-blind. Active intervention, placebo, and active controls not distinguished by look, taste, smell, packaging, or medication regimen.
Incomplete outcome data (attrition bias) | Unclear risk | Reported with withdrawals. (low risk)
| | | Per protocol analysis only. (unclear risk)
Selective reporting (reporting bias) | Low risk | Confirmatory design, sample size slightly smaller than planned, statistical power 80%, alpha 0.05
| | | Outcome VAS 0–10 converted to 100mm scale for data extraction
| | | Reported adverse events.
Other bias | Low risk | Diagnosis / assessment consistent with ACR criteria.
| | | Reported ethics committee approval.
| | | Reported clinical trials registration (ID-NCT00471055).
| | | Reported financial and in kind support.

Randall 2000

Methods | Randomised, double-blind, placebo control, 2 group crossover. Duration 12 weeks (2 × 1 week intervention, each followed by 5 week washout)
---|---
| Inclusion: persistent base of thumb pain (OA criteria not specified)
Interventions | Tradename not provided. *Urtica dioica* (stinging nettle).
| Placebo control: *Lamium album* (white dead nettle).
| Regimen: Whole leaf plucked from live plant, applied directly to skin of painful thumb, total contact with skin 30 seconds per day
Outcomes | Pain VAS 0–100 (0 indicates no pain, 100 indicates worst possible pain), pain (verbal 5 point scale, 0 indicates no pain, 5 indicates worst possible pain), HAQ-DI (higher score indicates more disability), analgesics, NSAID use, and sleep(scales not reported)
Notes | Results favour intervention, but mean improvement in HAQ-DI score does not exceed accepted minimal clinically important difference

Risk of bias

| Bias | Authors’ judgement | Support for judgement |
---|---|---|
Random sequence generation (selection bias) | Unclear risk | Described as randomised, method not reported.
Allocation concealment (selection bias) | Unclear risk | Allocation concealment not reported.
Blinding (performance bias and detection bias) | Unclear risk | Active intervention and placebo not distinguished by look, taste, or smell, but placebo validity and blinding may be compromised by stinging effect of active intervention
Incomplete outcome data (attrition bias) | Unclear risk | Reported withdrawals. (low risk)
| | | Per protocol analysis only. (unclear risk)
Selective reporting (reporting bias) | High risk | Exploratory study design; power, effect, and sample size not determined a priori. (unclear risk)
| | | Reported adverse events. (low risk)
| | | Included notes on unsatisfactory outcome measures. Most outcome data were reported as means only (not standard deviations) without confidence intervals. (unclear risk)
### Data for first and second periods were aggregated and insufficient for extraction from the first arm (up to crossover) only. (high risk) Aggregated data were extracted for the critical outcome of pain in the summary of findings tables. An error was identified during this data extraction. (unclear risk)

<table>
<thead>
<tr>
<th>Other bias</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria for diagnosis of OA not specified at baseline. Other arthritides are possible confounders. (high risk) Did not report ethical oversight.</td>
<td></td>
</tr>
</tbody>
</table>

---

### Randall 2008

**Methods**
Randomised, single blind, placebo control, 2 parallel groups. Duration 1 week intervention, plus 15 weeks follow-up

**Participants**
Randomised n=42, Completed intervention n=42, Completed follow up n=35. Mean age intervention group 65 yrs, control 67 yrs. M:F control 13:8, intervention 11:10. Inclusion: presumptive diagnosis of knee OA based on ACR criteria (Read diagnostic code/s for “knee pain” or “OA knee” in clinical records), aged 55–80 yrs, knee pain most days of the previous month, WOMAC pain subscale score of at least 4 at baseline

**Interventions**
Tradename not provided. *Urtica dioica* (stinging nettle). Placebo control: *Urtica galeopsifolia* (non-stinging nettle). Regimen: Whole leaf plucked from live plant, applied directly to skin of painful knee, total contact with knee 1 minute per day

**Outcomes**
WOMAC 0–4 (A, B, and C subscales; 24 items, 3 subscales, higher scores indicate worse deficits), pain at rest VAS 0–100, pain on walking VAS 0–100 (all VAS 0 indicates no pain, 100 indicates worst possible pain), patient global assessment of beneficial and adverse reactions, medication diary (scales not reported) Qualitative outcomes: focus groups discussions to explore participants’ attitudes and experiences of the trial

**Notes**
Results equivocal; stinging nettle not superior to placebo. Outcome scores for participants who returned poorly kept nettle plants did not differ significantly from group means

### Risk of bias

<table>
<thead>
<tr>
<th>Bias</th>
<th>Authors’ judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Low risk</td>
<td>Computer generated, block randomisation.</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Low risk</td>
<td>Adequate allocation concealment can be inferred. “Plants in serially numbered, opaque pots.”</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias)</td>
<td>Unclear risk</td>
<td>Active intervention and placebo not distinguishable by look, taste, or smell, but placebo validity and blinding may be compromised by stinging effect of active intervention</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias)</td>
<td>Unclear risk</td>
<td>Reported withdrawals. (low risk) Per protocol analysis only. (unclear risk)</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Unclear risk</td>
<td>Exploratory study design; power, effect, and sample size not determined a priori. (unclear risk) Reported adverse events. (low risk)</td>
</tr>
<tr>
<td>Other bias</td>
<td>Unclear risk</td>
<td>Diagnosis not consistent with ACR criteria. OA not distinguished from other causes of knee pain in older adults at baseline (presumptive diagnosis). Although OA is the most likely cause of knee pain in older adults, other arthritides are possible confounders. Reported ethics committee approval. (low risk)</td>
</tr>
</tbody>
</table>

---

### Soltanian 2010

**Methods**
Randomised, placebo control, double blind, single centre, crossover. Duration 6 weeks (2 × 3 week intervention, no washout)

**Participants**
Randomised n=42; intervention n=21, control n=21. Completed n=42; intervention n=21, control n=21. OA knee (EULAR criteria)

**Interventions**
Marhame-Mafasel: *Arnebia euchroma* and *Matricaria chamomilla* pomade. Placebo control: ingredients not reported, pomade. Regimen: 4.5g/day (1.5g TID) of pomade massaged firmly into skin until completely disappeared

**Outcomes**
WOMAC VAS 0–100 (24 items, 3 subscales, higher scores indicate worse deficits)
Results favour intervention. In all between and within-group comparisons, improvements in osteoarthritic pain, function, and stiffness were greater in people using Marhame-Mafasel over placebo. Improvement scores attributed to Marhame-Mafasel were somewhat greater in the second arm of the crossover, suggesting that there maybe a concurrent benefit from vigorous massage over time.

### Risk of bias

<table>
<thead>
<tr>
<th>Bias</th>
<th>Authors’ judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Low risk</td>
<td>Described as randomised. “Randomized according to a random number table.”</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Unclear risk</td>
<td>Allocation concealment not reported.</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias)</td>
<td>Unclear risk</td>
<td>Described as double blind. Active interventions not distinguished by look, taste, smell or packaging. (low risk)</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias)</td>
<td>Low risk</td>
<td>Reported no missing data, no withdrawals, no participants lost to follow up. (low risk)</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Unclear risk</td>
<td>Exploratory study design; power, effect, and sample size not determined a priori. (unclear risk) Reported adverse events. (low risk)</td>
</tr>
<tr>
<td>Other bias</td>
<td>Low risk</td>
<td>Diagnosis / assessment consistent with ACR criteria. (low risk) Report of ethics committee approval. (low risk) Reported that not all participants “completed” the intervention and that some participants displayed poor compliance with the intervention, but it is unclear how lack of compliance influenced results</td>
</tr>
</tbody>
</table>

### Wang 2012

**Methods**

Randomised, placebo control, double blind, single centre, 3 parallel groups (2 interventions). Duration 7 days

**Participants**

Randomised n=150; intervention A n=60, intervention B n=60, control n=30. Completed n=42; intervention n=21, control n=21. OA knee (Chinese Orthopaedic Association criteria), baseline pain >20mm on 100mm VAS

**Interventions**

Intervention A: topical patch containing Fufang Nanxing Zhitong Gao (FNZG) Chinese herbal mixture

Intervention B: topical patch containing Shangshi Jietong Gao (SJG) Chinese herbal mixture

Placebo control: topical patch made of acrylic, adhesive tape (no ingredients)

All patches applied to skin of right or left knee for 8 hours per day (overnight)

**Outcomes**

Pain VAS 0–100, WOMAC, Traditional Chinese Medicine Syndrome Questionnaire (TCMSQ)

**Notes**

Results favour FNZG patches.

### Risk of bias

<table>
<thead>
<tr>
<th>Bias</th>
<th>Authors’ judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Unclear risk</td>
<td>Described as randomised, method not reported.</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Unclear risk</td>
<td>Allocation reported as concealed, method not reported.</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias)</td>
<td>Low risk</td>
<td>Active interventions and placebo control not distinguished by look, taste, smell or packaging</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias)</td>
<td>Low risk</td>
<td>Reported withdrawals. Included per protocol and intention-to-treat analyses. Missing data replaced using last observation carried forward method</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Low risk</td>
<td>Confirmatory design, statistical power 80%, alpha 0.05.</td>
</tr>
</tbody>
</table>
All data reported as means and confidence intervals. Standard deviations calculated during data extraction. Reported adverse events.

<table>
<thead>
<tr>
<th>Bias</th>
<th>Authors’ judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other bias</td>
<td>Low risk</td>
<td>Diagnosis and assessment consistent with ACR criteria.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reported ethics committee approval, and clinical trials registration</td>
</tr>
</tbody>
</table>

**Widrig 2007**

**Methods**
Randomised, double blind, active control (ibuprofen 5% topical gel), 2 parallel groups, 20 centre study. Duration 3 weeks.

**Participants**

**Interventions**
A. Vogel Arnica Gel: Arnica montana (mountain arnica), tincture 50% v v.

**Outcomes**
Pain VAS 0–100 (0 indicates no pain, 100 indicates worst possible pain), hand algofunctional index 0–3 (10 items, higher score indicates worse function), tender joint count, morning stiffness intensity and duration, patient evaluation of efficacy, patient acceptance of treatment, physician evaluation of efficacy, acetaminophen use.

**Notes**
Arnica equally effective as ibuprofen on all primary and secondary measures. Post-intervention pain scores high with large standard deviation in both groups.

**Risk of bias**

<table>
<thead>
<tr>
<th>Bias</th>
<th>Authors’ judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Low risk</td>
<td>Described as randomised, method not reported.</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Low risk</td>
<td>Allocation concealment not reported.</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) All outcomes</td>
<td>Low risk</td>
<td>Active interventions not distinguished by look, taste, smellor packaging</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias) All outcomes</td>
<td>Low risk</td>
<td>Reported withdrawals. Included per protocol and intention-to-treat analyses</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Low risk</td>
<td>Confirmatory design, statistical power 80%, alpha 0.024. Non-inferiority hypothesis, MCID determined a priori at 12%.</td>
</tr>
<tr>
<td>Other bias</td>
<td>Low risk</td>
<td>Diagnosis / assessment consistent with ACR criteria.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reported ethics committee approval, and compliance with ICH GCP guidelines and Declaration of Helsinki</td>
</tr>
</tbody>
</table>

Unless otherwise stated, all oral medications are reported as total daily doses, which may have been administered in single or divided doses. Unless subscales are named, outcome measures (eg: WOMAC, HAQ, COAT) were used in entirety. Unless specified, all outcome measures were administered, scored, and scaled according to OARSI standards.

1. Reported compliance with ICH GCP guidelines (ICH 2004) anchored in European law, so adequate randomisation, allocation concealment, and blinding can be assumed.
### Characteristics of excluded studies [ordered by study ID]

<table>
<thead>
<tr>
<th>Study</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altman 1994</td>
<td>Intervention included extracted capsaicin, therefore not herbal as per WHO definition</td>
</tr>
<tr>
<td>Deal 1991</td>
<td>Intervention included extracted capsaicin, therefore not herbal as per WHO definition</td>
</tr>
<tr>
<td>Gemmell 2003</td>
<td>Intervention included capsaicin and menthol (extracted or synthetic), therefore not herbal as per WHO definition. Ingredients not listed in sufficient detail, therefore study is only repeatable using the proprietary product</td>
</tr>
<tr>
<td>Kielczynski 1997</td>
<td>Discussion paper.</td>
</tr>
<tr>
<td>Linsheng 1997</td>
<td>Not a randomised controlled trial. Case series.</td>
</tr>
<tr>
<td>Long 2001</td>
<td>Review paper.</td>
</tr>
<tr>
<td>McCarthy 1992</td>
<td>Intervention included extracted capsaicin, therefore not herbal as per WHO definition</td>
</tr>
<tr>
<td>McCleane 2000</td>
<td>Intervention included extracted capsaicin, therefore not herbal as per WHO definition</td>
</tr>
<tr>
<td>McKay 2003</td>
<td>Intervention included capsaicin and menthol (extracted or synthetic), therefore not herbal as per WHO definition. Ingredients not listed in sufficient detail, therefore study is only repeatable using the proprietary product</td>
</tr>
<tr>
<td>Rayburn 2009</td>
<td>Not a randomised controlled trial.</td>
</tr>
<tr>
<td>Sagar 1988</td>
<td>Not a randomised controlled trial.</td>
</tr>
<tr>
<td>Saley 1987</td>
<td>Not a randomised controlled trial. Case series.</td>
</tr>
<tr>
<td>Schnitzer 1994</td>
<td>Intervention included extracted capsaicin, therefore not herbal as per WHO definition</td>
</tr>
<tr>
<td>Smith 2011</td>
<td>Intervention included tannic acid (extracted or synthetic), therefore not herbal as per WHO definition</td>
</tr>
<tr>
<td>Wadmap 2006</td>
<td>Not a randomised controlled trial (observational study).</td>
</tr>
<tr>
<td>Yuelong 2011</td>
<td>Not a randomised controlled trial. Protocol for a randomised controlled trial only</td>
</tr>
</tbody>
</table>
### Characteristics of studies awaiting assessment [ordered by study ID]

<table>
<thead>
<tr>
<th>Study</th>
<th>Methods</th>
<th>Participants</th>
<th>Interventions</th>
<th>Outcomes</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zahmatkash 2011</td>
<td>RCT</td>
<td>n=92</td>
<td>Intervention: herbal ointment containing cinnamon, ginger, mastic (Saghez) and sesame oil Active control: salicylate ointment.</td>
<td>Pain VAS 0–100, stiffness VAS 0–100, limited motion VAS 0–100</td>
<td>Head to head comparison, non-inferiority hypothesis. Abstract only available. <em>Pakistan Journal of Biological Science</em> is not indexed. Full manuscript sought in hard copy via inter-library loan</td>
</tr>
<tr>
<td>Zhong 2006</td>
<td>RCT</td>
<td>n=88 (intervention n=44, control n=44)</td>
<td>Intervention: Bushen Qhuan Tongluo herbs by orally or externally washing Bushen Qhuan. Tongluo: Hutaorou (12 g), Buguzhi (12 g), Chaozuzhong (12 g), Shudi (15 g), Dahuixiang (9 g), Laoshiteng (15 g), Zhichuanwu (9 g), Sanqi (6 g), Wugong (3 g), Jixieteng (15 g). The prescription for external washing: Tuogucao (40 g), Danggui (15 g), Sumu (15 g), Shengduhuang (15 g), Shengnanxing (10 g), Ruxiang (10 g), Meyao (10 g), Bingpian (3 g). Topical administration: The medicine that dissolved in 500 mL of 100 degreesC boiled water was adopted to wash both knees while the temperature down to 35 degreesC one dose upon a time and twice a day) Patients in the control group were given sulfated glucosamine (Weiguli Capsule. Each capsule contains 314 mg of sulfated glucosamine crystal, which is equal to 250 mg of sulfated glucosamine) two capsules a time and 3 times a day as well as piroxicam (Yantong Xikang Pill) once a day and 20 mg each time. Patients in both groups were administrated for 12 weeks.</td>
<td>WOMAC</td>
<td>Unable to distinguish oral administration intervention group from topical administration intervention group results from abstract alone. Abstract only available. <em>Chinese Journal of Clinical Rehabilitation</em> not indexed. Full manuscript sought in hard copy via inter-library loan</td>
</tr>
</tbody>
</table>
## Characteristics of ongoing studies [ordered by study ID]

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Trial name or title</th>
<th>Methods</th>
<th>Participants</th>
<th>Interventions</th>
<th>Outcomes</th>
<th>Starting date</th>
<th>Contact information</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Youxin 2012</td>
<td>Rehabilitation protocol of the Traditional Chinese Medicine on patients with dyskinesia of the knee osteoarthritis: A randomized controlled trial study in community</td>
<td>Randomised, controlled, parallel 2-group trial.</td>
<td>Community dwelling male and female adults aged 40–75 years, with knee OA (Chinese Orthopaedic Association criteria); n = 722 (intervention n= 361, control n = 361)</td>
<td>Foundation treatment of Chinese medicinal herb washing and traditional exercises training method, plus blood-letting puncture, acupuncture, massage, and analgesics</td>
<td>Pain, swelling, knee range of motion, muscle strength, average walking distance, stair climbing and descent, activities of daily living, analgesic use, quality of life, safety, health economic evaluation, global effect</td>
<td>Unknown, Ethics committee approval from 27/09/2012.</td>
<td>Su Youxin: No. 1, HuATuo road, Shangjie town, Minhou county, Fujian, Fuzhou, China; +86 1330 502 1666; <a href="mailto:suyouxin777@hotmail.com">suyouxin777@hotmail.com</a></td>
<td>Prospective registration, ongoing clinical trial.</td>
</tr>
</tbody>
</table>